

# Average Consensus on General Strongly Connected Digraphs

Kai Cai and Hideaki Ishii

## Abstract

We study the average consensus problem of multi-agent systems for general network topologies with unidirectional information flow. We propose two (linear) distributed algorithms, deterministic and gossip, respectively for the cases where the inter-agent communication is synchronous and asynchronous. Our contribution is that in both cases, the developed algorithms guarantee state averaging on arbitrary strongly connected digraphs; in particular, this graphical condition does not require that the network be balanced or symmetric, thereby extending many previous results in the literature. The key novelty of our approach is to augment an additional variable for each agent, called “surplus”, whose function is to locally record individual state updates. For convergence analysis, we employ graph-theoretic and nonnegative matrix tools, with the eigenvalue perturbation theory playing a crucial role.

## I. INTRODUCTION

This paper presents a new approach to the design of distributed algorithms for *average consensus*: that is, a system of networked agents reaches an agreement on the average value of their initial states, through merely local interaction among peers. The approach enables multi-agent systems to achieve average consensus on arbitrary strongly connected network topologies with unidirectional information flow, where the state sum of the agents need not stay put as time evolves.

There has been an extensive literature addressing multi-agent consensus problems. Many fundamental distributed algorithms (developed in, e.g., [1]–[5]) are of the *synchronous* type: At an arbitrary time, individual agents are assumed to sense and/or communicate with all the neighbors, and then simultaneously execute their local protocols. In particular, Olfati-Saber and Murray [3] studied algorithms of such type to achieve average consensus on static digraphs, and justified that a *balanced* and *strongly connected*

The authors are with the Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, 4259-J2-54, Nagatsuta-cho, Midori-ku, Yokohama 226-8502, Japan. Phone/Fax: +81-45-924-5371. Emails: caikai@sc.dis.titech.ac.jp, ishii@dis.titech.ac.jp. This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology in Japan under Grant-in-Aid for Scientific Research, No. 21760323.

topology is necessary and sufficient to guarantee convergence. More recently in [6], Boyd et al. proposed a compelling “gossip” algorithm, which provides an asynchronous approach to treat average consensus. Specifically, the algorithm assumes that at each time instant, exactly one agent wakes up, contacts only one of its neighbors selected at random, and then these two agents average out their states. The graph model that the algorithm bases is undirected (or symmetric), and average consensus is ensured as long as the topology is *connected*. Since then, the gossip approach has been widely adopted [7]–[9] in tackling average consensus on undirected graphs, with additional constraints on quantized information flow; see also [10] for related distributed computation problems in search engines.

In this paper, and its conference precursor [11], we study the average consensus problem under both synchronous and asynchronous setups, as in [3] and [6]. In both cases, we propose a novel type of (linear) distributed algorithms, which can be seen as extensions of the corresponding algorithms in [3] and [6]; and we prove that these new algorithms guarantee state averaging on arbitrary strongly connected digraphs, therefore generalizing the graphical conditions derived in [3] and [6]. We note that digraph models have been studied extensively in the consensus literature [3]–[5], and considered to be generally more economical for information exchange than their undirected counterpart (refer to [3] for more detailed motivation of using digraphs). Our underlying (theoretic) interest in this paper is to generalize the connectivity conditions on digraphs for average consensus.

The primary challenge of average consensus on arbitrary strongly connected digraphs lies in that the state sum of agents need not be preserved, thereby causing shifts in the average. We note that there are a few efforts in the literature having addressed this issue. In [12], an auxiliary variable is associated to each agent and a linear broadcast gossip algorithm is proposed; however, the convergence of that algorithm is not proved, and remarked to be difficult. References [13], [14] also use extra variables, and a nonlinear (division involved) algorithm is designed and proved to achieve state averaging on non-balanced digraphs. The idea is based on computing the stationary distribution for the Markov chain characterized by the agent network, and is thus quite different from consensus-type algorithms [3], [6]. In [1, Section 7.4], the load balancing problem is tackled in which inter-processor communication is asynchronous and with bounded delay. The underlying topology is assumed undirected; owing to asynchronism and delay, however, the total amount of loads at processors is not invariant. A switched linear algorithm is proposed to achieve load balancing in this scenario, the rules of which rely on however bidirectional communication. In addition, a different and interesting approach is presented in [15]: Given a general strongly connected digraph, find a corresponding doubly stochastic matrix (which, when used as a distributed updating scheme, guarantees state averaging [4]). An algorithm is designed to achieve this goal by adding self-loop edges with proper

weights to balance flow-in and -out information. Finally, time-varying state sum caused by packet loss or communication failure is considered in [16], [17], and the deviation from the initial average is analyzed.

We develop a new approach to handle the problem that the state sum of agents need not be preserved. Similar to [12], we also augment an additional variable for each agent, which we call “surplus”; different from [13], [14], the function of surplus variables is to record every state change of the associated agent. Thus, in effect, these variables collectively maintain the information of the average shift amount.<sup>1</sup> Using this novel idea, our main contribution is the design of linear algorithms (without switching) to achieve average consensus on general strongly connected digraphs, in contrast with the types of algorithms designed in [13], [14] and [1, Section 7.4]. Also, linearity allows us to employ certain matrix tools in analysis, which are very different from the proof methods used in [13], [14] and [1, Section 7.4]. Moreover, our technical contribution in this paper is the demonstration of matrix perturbation tools (including eigenvalue perturbation, optimal matching distance, and Bauer-Fike Theorem [20]–[22]) in analyzing convergence properties, which seems unexplored in the consensus literature.

Our idea of adding surpluses is indeed a continuation of our own previous work in [23], where the original surplus-based approach is proposed to tackle *quantized* average consensus on general digraphs. There we developed a quantized (thus nonlinear) averaging algorithm, and the convergence analysis is based on finite Markov chains. By contrast, the algorithms designed in the present paper are linear, and hence the convergence can be characterized by the spectral properties of the associated matrices. On the other hand, our averaging algorithms differ also from those basic ones [3], [6] in that the associated matrices contain negative entries. Consequently for our analysis tools, besides the usual nonnegative matrix theory and algebraic graph theory, the matrix perturbation theory is found instrumental.

The paper is organized as follows. Section II formulates distributed average consensus problems in both synchronous and asynchronous setups. Sections III and IV present the respective solution algorithms, which are rigorously proved to guarantee state averaging on general strongly connected digraphs. Further, Section V explores certain special topologies that lead us to specialized results, and Section VI provides a set of numerical examples for demonstration. Finally, Section VII states our conclusions.

*Notation.* Let  $\mathbf{1} := [1 \cdots 1]^T \in \mathbb{R}^n$  be the vector of all ones. For a complex number  $\lambda$ , denote its real part by  $\text{Re}(\lambda)$ , imaginary part by  $\text{Im}(\lambda)$ , conjugate by  $\bar{\lambda}$ , and modulus by  $|\lambda|$ . For a set  $\mathcal{S}$ , denote its

<sup>1</sup>The method of augmenting auxiliary variables is also found in [18] and [19], as *predictors* estimating future states and *shift registers* storing past states respectively, in order to accelerate consensus speed. How the predictors or registers are used in these references is, however, very different from our usage of surpluses.

cardinality by  $\text{card}(\mathcal{S})$ . Given a real number  $x$ ,  $\lfloor x \rfloor$  is the largest integer smaller than or equal to  $x$ , and  $\lceil x \rceil$  is the smallest integer larger than or equal to  $x$ . Given a matrix  $M$ ,  $|M|$  denotes its determinant; the spectrum  $\sigma(M)$  is the set of its eigenvalues; the spectral radius  $\rho(M)$  is the maximum modulus of its eigenvalues. In addition,  $\|\cdot\|_2$  and  $\|\cdot\|_\infty$  denote the 2-norm and infinity norm of a vector/matrix.

## II. PROBLEM FORMULATION

Given a network of  $n$  ( $> 1$ ) agents, we model its interconnection structure by a digraph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ : Each *node* in  $\mathcal{V} = \{1, \dots, n\}$  stands for an agent, and each directed *edge*  $(j, i)$  in  $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$  denotes that agent  $j$  communicates to agent  $i$  (namely, the information flow is from  $j$  to  $i$ ). Selfloop edges are not allowed, i.e.,  $(i, i) \notin \mathcal{E}$ . In  $\mathcal{G}$  a node  $i$  is *reachable* from a node  $j$  if there exists a path from  $j$  to  $i$  which respects the direction of the edges. We say  $\mathcal{G}$  is *strongly connected* if every node is reachable from every other node. A *closed strong component* of  $\mathcal{G}$  is a maximal set of nodes whose corresponding subdigraph is strongly connected and closed (i.e., no node inside the subdigraph is reachable from any node outside). Also a node  $i$  is called *globally reachable* if every other node is reachable from  $i$ .

At time  $k \in \mathbb{Z}_+$  (nonnegative integers) each node  $i \in \mathcal{V}$  has a scalar state  $x_i(k) \in \mathbb{R}$ ; the aggregate state is denoted by  $x(k) = [x_1(k) \cdots x_n(k)]^T \in \mathbb{R}^n$ . The *average consensus problem* aims at designing distributed algorithms, where individual nodes update their states using only the local information of their neighboring nodes in the digraph  $\mathcal{G}$  such that all  $x_i(k)$  eventually converge to the initial average  $x_a := \mathbf{1}^T x(0)/n$ . To achieve state averaging on general digraphs, the main difficulty is that the state sum  $\mathbf{1}^T x$  need not remain invariant, which can result in losing track of the initial average  $x_a$ . To deal with this problem, we propose associating to each node  $i$  an additional variable  $s_i(k) \in \mathbb{R}$ , called *surplus*; write  $s(k) = [s_1(k) \cdots s_n(k)]^T \in \mathbb{R}^n$  and set  $s(0) = 0$ . The function of surplus is to locally record the state changes of individual nodes such that  $\mathbf{1}^T(x(k) + s(k)) = \mathbf{1}^T x(0)$  for all time  $k$ ; in other words, surplus keeps the quantity  $\mathbf{1}^T(x + s)$  constant over time.

In the first part of this paper, we consider synchronous networks as in [3]: At each time, every node communicates with all of its neighbors simultaneously, and then makes a corresponding update.

*Definition 1.* A network of agents is said to achieve *average consensus* if for every initial condition  $(x(0), s(0) = 0)$ , it holds that  $(x(k), s(k)) \rightarrow (x_a \mathbf{1}, 0)$  as  $k \rightarrow \infty$ .

*Problem 1.* Design a distributed algorithm such that agents achieve average consensus on general digraphs.

To solve this problem, we will propose in Section III a surplus-based distributed algorithm, under which we justify that average consensus is achieved for general strongly connected digraphs.

In the second part, we consider the setup of asynchronous networks as in [6]. Specifically, communication among nodes is by means of *gossip*: At each time, exactly one edge  $(j, i) \in \mathcal{E}$  is activated at random, independently from all earlier instants and with a time-invariant, strictly positive probability  $p_{ij} \in (0, 1)$  such that  $\sum_{(j,i) \in \mathcal{E}} p_{ij} = 1$ . Along this activated edge, node  $j$  sends its state and surplus to node  $i$ , while node  $i$  receives the information and makes a corresponding update.

*Definition 2.* A network of agents is said to achieve

- (i) *mean-square average consensus* if for every initial condition  $(x(0), s(0) = 0)$ , it holds that  $E[\|x(k) - x_a \mathbf{1}\|_2^2] \rightarrow 0$  and  $E[\|s(k)\|_2^2] \rightarrow 0$  as  $k \rightarrow \infty$ ;
- (ii) *almost sure average consensus* if for every initial condition  $(x(0), s(0) = 0)$ , it holds that  $(x(k), s(k)) \rightarrow (x_a \mathbf{1}, 0)$  as  $k \rightarrow \infty$  with probability one.

As defined, the mean-square convergence is concerned with the second moments of the state and surplus processes, whereas the almost sure convergence is with respect to the corresponding sample paths. It should be noted that in general there is no implication between these two convergence notions (e.g., [24, Section 7.2]).

*Problem 2.* Design a distributed algorithm such that agents achieve mean-square and/or almost sure average consensus on general digraphs.

For this problem, we will propose in Section IV a surplus-based gossip algorithm, under which we justify that both mean-square and almost sure average consensus can be achieved for general strongly connected digraphs.

### III. AVERAGING IN SYNCHRONOUS NETWORKS

This section solves Problem 1. First we present a (discrete-time) distributed algorithm based on surplus, which may be seen as an extension of the standard consensus algorithms in the literature [1]–[5]. Then we prove convergence to average consensus for general strongly connected digraphs.

#### A. Algorithm Description

Consider a system of  $n$  agents represented by a digraph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ . For each node  $i \in \mathcal{V}$ , let  $\mathcal{N}_i^+ := \{j \in \mathcal{V} : (j, i) \in \mathcal{E}\}$  denote the set of its “in-neighbors”, and  $\mathcal{N}_i^- := \{h \in \mathcal{V} : (i, h) \in \mathcal{E}\}$  the set of its “out-neighbors”. Note that  $\mathcal{N}_i^+ \neq \mathcal{N}_i^-$  in general; and  $i \notin \mathcal{N}_i^+$  or  $\mathcal{N}_i^-$ , for selfloop edges do not exist. There are three operations that every node  $i$  performs at time  $k \in \mathbb{Z}_+$ . First, node  $i$  sends its state information  $x_i(k)$  and weighted surplus  $b_{ih}s_i(k)$  to each out-neighbor  $h \in \mathcal{N}_i^-$ ; here the *sending weight*

$b_{ih}$  is assumed to satisfy that  $b_{ih} \in (0, 1)$  if  $h \in \mathcal{N}_i^-$ ,  $b_{ih} = 0$  if  $h \in \mathcal{V} - \mathcal{N}_i^-$ , and  $\sum_{h \in \mathcal{N}_i^-} b_{ih} < 1$ . Second, node  $i$  receives state information  $x_j(k)$  and weighted surplus  $b_{ji}s_j(k)$  from each in-neighbor  $j \in \mathcal{N}_i^+$ . Third, node  $i$  updates its own state  $x_i(k)$  and surplus  $s_i(k)$  as follows:

$$x_i(k+1) = x_i(k) + \sum_{j \in \mathcal{N}_i^+} a_{ij}(x_j(k) - x_i(k)) + \epsilon s_i(k), \quad (1)$$

$$s_i(k+1) = \left( (1 - \sum_{h \in \mathcal{N}_i^-} b_{ih})s_i(k) + \sum_{j \in \mathcal{N}_i^+} b_{ji}s_j(k) \right) - (x_i(k+1) - x_i(k)), \quad (2)$$

where the *updating weight*  $a_{ij}$  is assumed to satisfy that  $a_{ij} \in (0, 1)$  if  $j \in \mathcal{N}_i^+$ ,  $a_{ij} = 0$  if  $j \in \mathcal{V} - \mathcal{N}_i^+$ , and  $\sum_{j \in \mathcal{N}_i^+} a_{ij} < 1$ ; in addition, the parameter  $\epsilon$  is a positive number which specifies the amount of surplus used to update the state.

We discuss the implementation of the above protocol in applications like sensor networks. Let  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  represent a network sensor nodes. Our protocol deals particularly with scenarios where (i) sensors have different communication ranges owing possibly to distinct types or power supplies; (ii) communication is by means of broadcasting (e.g., [12]) which again might have different ranges; and (iii) strategy of random geographic routing is used for efficient and robust node value aggregation in one direction [13], [14]. In these scenarios, information flow among sensors is typically directed. A concrete example is using sensor networks for monitoring geological areas (e.g., volcanic activities), where sensors are fixed at certain locations. At the time of setting them up, the sensors may be given different transmission power for saving energy (such sensors must run for a long time) or owing to geological reasons. Once the power is fixed, the neighbors (and their IDs) can be known to each sensor. Thus, digraphs can arise in *static* sensor networks where the neighbors can be fixed and known. To implement states and surpluses, we see from (1), (2) that they are ordinary variables locally stored, updated, and exchanged; thus they may be implemented by allocating memories in sensors. For the parameter  $\epsilon$ , we will see that it plays a crucial role in the convergence of our algorithm; however,  $\epsilon$  must be chosen sufficiently small, and a valid bound for its value involves non-local information of the digraph. The latter constraint (in bounding a parameter) is often found in consensus algorithms involving more than one variable [5], [25], [26]. One may overcome this by computing a valid bound offline, and notify that  $\epsilon$  value to every node.

Now let the *adjacency matrix*  $A$  of the digraph  $\mathcal{G}$  be given by  $A := [a_{ij}] \in \mathbb{R}^{n \times n}$ , where the entries are the updating weights. Then the *Laplacian matrix*  $L$  is defined as  $L := D - A$ , where  $D = \text{diag}(d_1, \dots, d_n)$  with  $d_i = \sum_{j=1}^n a_{ij}$ . Thus  $L$  has nonnegative diagonal entries, nonpositive off-diagonal entries, and zero row sums. Then the matrix  $I - L$  is nonnegative (by  $\sum_{j \in \mathcal{N}_i^+} a_{ij} < 1$ ), and every row sums up to one; namely  $I - L$  is *row stochastic*. Also let  $B := [b_{ih}]^T \in \mathbb{R}^{n \times n}$ , where the entries are the sending

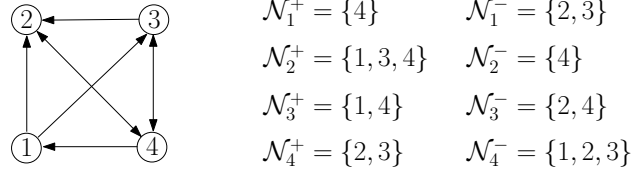


Fig. 1. Illustrating example of four agents: communication topology and neighbor sets.

weights (note that the transpose in the notation is needed because  $h \in \mathcal{N}_i^-$  for  $b_{ih}$ ). Define the matrix  $S := (I - \tilde{D}) + B$ , where  $\tilde{D} = \text{diag}(\tilde{d}_1, \dots, \tilde{d}_n)$  with  $\tilde{d}_i = \sum_{h=1}^n b_{ih}$ . Then  $S$  is nonnegative (by  $\sum_{h \in \mathcal{N}_i^-} b_{ih} < 1$ ), and every column sums up to one; i.e.,  $S$  is *column stochastic*. As can be observed from (2), the matrix  $S$  captures the part of update induced by sending and receiving surplus.

With the above matrices, the iterations (1) and (2) can be written in a matrix form as

$$\begin{bmatrix} x(k+1) \\ s(k+1) \end{bmatrix} = M \begin{bmatrix} x(k) \\ s(k) \end{bmatrix}, \quad \text{where } M := \begin{bmatrix} I - L & \epsilon I \\ L & S - \epsilon I \end{bmatrix} \in \mathbb{R}^{2n \times 2n}. \quad (3)$$

Notice that (i) the matrix  $M$  has negative entries due to the presence of the Laplacian matrix  $L$  in the  $(2, 1)$ -block; (ii) the column sums of  $M$  are equal to one, which implies that the quantity  $x(k) + s(k)$  is a constant for all  $k \geq 0$ ; and (iii) the state evolution specified by the  $(1, 1)$ -block of  $M$ , i.e.,

$$x(k+1) = (I - L)x(k), \quad (4)$$

is that of the *standard* consensus algorithm studied in the literature (e.g., [1], [3], [4]). We will henceforth refer to (3) as the deterministic algorithm, and analyze its convergence properties in the next subsection.

*Example 1.* For an illustration of the algorithm (3), consider a network of four nodes with neighbor sets shown in Fig. 1. Fixing  $i \in [1, 4]$ , let  $a_{ij} = 1/(\text{card}(\mathcal{N}_i^+) + 1)$  for every  $j \in \mathcal{N}_i^+$  and  $b_{ih} = 1/(\text{card}(\mathcal{N}_i^-) + 1)$  for every  $h \in \mathcal{N}_i^-$ . Then the matrix  $M$  of this example is given by

$$M = \left[ \begin{array}{cccc|cccc} 1/2 & 0 & 0 & 1/2 & \epsilon & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 & \epsilon & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 & 0 & \epsilon & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & \epsilon \\ \hline 1/2 & 0 & 0 & -1/2 & 1/3 - \epsilon & 0 & 0 & 1/4 \\ -1/4 & 3/4 & -1/4 & -1/4 & 1/3 & 1/2 - \epsilon & 1/3 & 1/4 \\ -1/3 & 0 & 2/3 & -1/3 & 1/3 & 0 & 1/3 - \epsilon & 1/4 \\ 0 & -1/3 & -1/3 & 2/3 & 0 & 1/2 & 1/3 & 1/4 - \epsilon \end{array} \right].$$

We see that  $M$  has negative entries, and every column sums up to one.

### B. Convergence Result

The following is a graphical characterization for the deterministic algorithm (3) to achieve average consensus. The proof is deferred to Section III-C.

*Theorem 1.* Using the deterministic algorithm (3) with the parameter  $\epsilon > 0$  sufficiently small, the agents achieve average consensus if and only if the digraph  $\mathcal{G}$  is strongly connected.

Without augmenting surplus variables, it is well known [3] that a necessary and sufficient graphical condition for state averaging is that the digraph  $\mathcal{G}$  is both strongly connected and *balanced*<sup>2</sup>. A balanced structure can be restrictive because when all the weights  $a_{ij}$  are identical, it requires the number of incoming and outgoing edges at each node in the digraph to be the same. By contrast, our algorithm (3) ensures average consensus for arbitrary strongly connected digraphs (including those non-balanced).

A surplus-based averaging algorithm was initially proposed in [23] for a quantized consensus problem. It guarantees that the integer-valued states converge to either  $\lfloor x_a \rfloor$  or  $\lceil x_a \rceil$ ; however, the steady-state surpluses are nonzero in general. There, the set of states and surpluses is finite, and thus arguments of finite Markov chain type are employed in the proof. Distinct from [23], with the algorithm (3) the states converge to the exact average  $x_a$  and the steady-state surpluses are zero. Moreover, since the algorithm (3) is linear, its convergence can be analyzed using tools from matrix theory, as detailed below. This last linearity point is also in contrast with the division involved algorithm designed in [13], [14].

The choice of the parameter  $\epsilon$  depends on the graph structure and the number of agents. In the following, we present an upper bound on  $\epsilon$  for general networks.

*Proposition 1.* Suppose that the digraph  $\mathcal{G}$  is strongly connected. The deterministic algorithm (3) achieves average consensus if the parameter  $\epsilon$  satisfies  $\epsilon \in (0, \bar{\epsilon}^{(d)})$ , where

$$\bar{\epsilon}^{(d)} := \frac{1}{(20 + 8n)^n} (1 - |\lambda_3|)^n, \quad \text{with } \lambda_3 \text{ the third largest eigenvalue of } M \text{ by setting } \epsilon = 0. \quad (5)$$

The proof of Proposition 1 is presented in Section III-D, which employs a fact from matrix perturbation theory (e.g., [21], [22]) relating the size of  $\epsilon$  to the distance between perturbed and unperturbed eigenvalues. Also, we will stress that this proof is based on that of Theorem 1. The above bound  $\bar{\epsilon}^{(d)}$  ensures average consensus for arbitrary strongly connected topologies. Due to the power  $n$ , however, the bound is rather conservative. This power is unavoidable for any perturbation bound result with respect to general matrices, as is well known in matrix perturbation literature [21], [22]. In Section V, we will

<sup>2</sup>A digraph  $\mathcal{G}$  with its adjacency matrix  $A = [a_{ij}]$  is balanced if  $\sum_{j=1}^n a_{ij} = \sum_{j=1}^n a_{ji}$  for all  $i$ . Equivalently, the system matrix  $I - L$  of the standard consensus algorithm (4) is both row and column stochastic [3], [4].

exploit structures of some special topologies, which yield less conservative bounds on  $\epsilon$ . Also, we see that the bound in (5) involves  $\lambda_3$ , the second largest eigenvalue of either  $I - L$  or  $S$  (matrix  $M$  is block-diagonal when  $\epsilon = 0$ ). This infers that, in order to bound  $\epsilon$ , we need to know the structure of the agent network. Such a requirement when bounding some parameters in consensus algorithms, unfortunately, seems to be not unusual [5], [25], [26].

### C. Proof of Theorem 1

We present the proof of Theorem 1. First, we state a necessary and sufficient condition for average consensus in terms of the spectrum of the matrix  $M$ .

*Proposition 2.* The deterministic algorithm (3) achieves average consensus if and only if 1 is a simple eigenvalue of  $M$ , and all other eigenvalues have moduli smaller than one.

*Proof.* (Sufficiency) Since every column of  $M$  sums up to one, 1 is an eigenvalue of  $M$  and  $[\mathbf{1}^T \ \mathbf{1}^T]^T$  is a corresponding left eigenvector. Note also that  $M[\mathbf{1}^T \ 0]^T = [\mathbf{1}^T \ 0]^T$ ; so  $[\mathbf{1}^T \ 0]^T \in \mathbb{R}^{2n}$  is a right eigenvector corresponding to the eigenvalue 1. Write  $M$  in Jordan canonical form as

$$M = VJV^{-1} = \begin{bmatrix} y_1 & \cdots & y_{2n} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & J' \end{bmatrix} \begin{bmatrix} z_1^T \\ \vdots \\ z_{2n}^T \end{bmatrix},$$

where  $y_i, z_i \in \mathbb{C}^{2n}$ ,  $i \in [1, 2n]$ , are respectively the (generalized) right and left eigenvectors of  $M$ ; and  $J' \in \mathbb{C}^{(2n-1) \times (2n-1)}$  contains the Jordan block matrices corresponding to those eigenvalues with moduli smaller than one. For the eigenvalue 1 choose  $y_1 = [\mathbf{1}^T \ 0]^T$  and  $z_1 = (1/n)[\mathbf{1}^T \ \mathbf{1}^T]^T$ ; thus  $z_1^T y_1 = 1$ . Now the  $k$ th power of  $M$  is

$$M^k = VJ^kV^{-1} = V \begin{bmatrix} 1 & 0 \\ 0 & (J')^k \end{bmatrix} V^{-1} \rightarrow y_1 z_1^T = \begin{bmatrix} \frac{1}{n} \mathbf{1} \mathbf{1}^T & \frac{1}{n} \mathbf{1} \mathbf{1}^T \\ 0 & 0 \end{bmatrix}, \quad \text{as } k \rightarrow \infty.$$

Therefore

$$\begin{bmatrix} x(k) \\ s(k) \end{bmatrix} = M^k \begin{bmatrix} x(0) \\ s(0) \end{bmatrix} \rightarrow \begin{bmatrix} \frac{1}{n} \mathbf{1} \mathbf{1}^T & \frac{1}{n} \mathbf{1} \mathbf{1}^T \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x(0) \\ s(0) \end{bmatrix} = \begin{bmatrix} \frac{1}{n} \mathbf{1} \mathbf{1}^T x(0) \\ 0 \end{bmatrix} = \begin{bmatrix} x_a \mathbf{1} \\ 0 \end{bmatrix}, \quad \text{as } k \rightarrow \infty.$$

(Necessity) First we claim that the eigenvalue 1 of  $M$  is always simple. Suppose on the contrary that the algebraic multiplicity of 1 equals two. The corresponding geometric multiplicity, however, equals one; this is checked by verifying  $\text{rank}(M - I) = 2n - 1$ . Thus there exists a generalized right eigenvector

$u = [u_1^T \ u_2^T]^T \in \mathbb{R}^{2n}$  such that  $(M - I)^2 u = 0$ , and  $(M - I)u$  is a right eigenvector with respect to the eigenvalue 1. Since  $[\mathbf{1}^T \ 0]^T$  is also a right eigenvector corresponding to the eigenvalue 1, it must hold:

$$\begin{aligned} (M - I)u &= c[\mathbf{1}^T \ 0]^T, \quad \text{for some scalar } c \neq 0 \\ \Rightarrow \begin{bmatrix} -L & \epsilon I \\ L & S - I - \epsilon I \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} &= c \begin{bmatrix} \mathbf{1} \\ 0 \end{bmatrix} \\ \Rightarrow \begin{cases} -Lu_1 + \epsilon u_2 = c\mathbf{1} \\ Lu_1 + (S - I)u_2 - \epsilon u_2 = 0 \end{cases} \\ \Rightarrow (S - I)u_2 &= c\mathbf{1}. \end{aligned}$$

One may verify that  $\text{rank}(S - I) = n - 1$  but  $\text{rank}([S - I \ c\mathbf{1}]) = n$ . Hence there is no solution for  $u_2$ , which in turn implies that the generalized right eigenvector  $u$  cannot exist. This proves our claim.

Now suppose that there is an eigenvalue  $\lambda$  of  $M$  such that  $\lambda \neq 1$  and  $|\lambda| \geq 1$ . But this immediately implies that  $\lim_{k \rightarrow \infty} M^k$  does not exist [4]. Therefore, average consensus cannot be achieved. ■

Next, we introduce an important result from matrix perturbation theory (e.g., [20, Chapter 2]), which is found crucial in analyzing the spectral properties of the matrix  $M$  in (3). The proof of this result can be found in [20, Sections 2.8 and 2.10]. An eigenvalue of a matrix is said *semi-simple* if its algebraic multiplicity is equal to its geometric multiplicity.

*Lemma 1.* Consider an  $n \times n$  matrix  $W(\epsilon)$  which depends smoothly on a real parameter  $\epsilon \geq 0$ . Fix  $l \in [1, n]$ ; let  $\lambda_1 = \dots = \lambda_l$  be a semi-simple eigenvalue of  $W(0)$ , with (linearly independent) right eigenvectors  $y_1, \dots, y_l$  and (linearly independent) left eigenvectors  $z_1, \dots, z_l$  such that

$$\begin{bmatrix} z_1^T \\ \vdots \\ z_l^T \end{bmatrix} \begin{bmatrix} y_1 & \dots & y_l \end{bmatrix} = I.$$

Consider a small  $\epsilon > 0$ , and denote by  $\lambda_i(\epsilon)$  the eigenvalues of  $W(\epsilon)$  corresponding to  $\lambda_i$ ,  $i \in [1, l]$ . Then the derivatives  $d\lambda_i(\epsilon)/d\epsilon|_{\epsilon=0}$  exist, and they are the eigenvalues of the following  $l \times l$  matrix:

$$\begin{bmatrix} z_1^T \dot{W} y_1 & \dots & z_1^T \dot{W} y_l \\ \vdots & & \vdots \\ z_l^T \dot{W} y_1 & \dots & z_l^T \dot{W} y_l \end{bmatrix}, \quad \text{where } \dot{W} := dW(\epsilon)/d\epsilon|_{\epsilon=0}. \quad (6)$$

Now we are ready to prove Theorem 1. The necessity argument follows from the one for [23, Theorem 2]; indeed, the class of strongly connected digraphs characterizes the existence of a distributed

algorithm that can solve average consensus. For the sufficiency part, let

$$M_0 := \begin{bmatrix} I - L & 0 \\ L & S \end{bmatrix} \quad \text{and} \quad F := \begin{bmatrix} 0 & I \\ 0 & -I \end{bmatrix}. \quad (7)$$

Then  $M = M_0 + \epsilon F$ , and we view  $M$  as being obtained by “perturbing”  $M_0$  via the term  $\epsilon F$ . Also, it is clear that  $M$  depends smoothly on  $\epsilon$ . Concretely, we will first show that the eigenvalues  $\lambda_i$  of the unperturbed matrix  $M_0$  satisfy

$$1 = \lambda_1 = \lambda_2 > |\lambda_3| \geq \dots \geq |\lambda_{2n}|. \quad (8)$$

Then using Lemma 1 we will establish that after a small perturbation  $\epsilon F$ , the obtained matrix  $M$  has only a simple eigenvalue 1 and all other eigenvalues have moduli smaller than one. This is the characteristic part of our proof. Finally, it follows from Proposition 2 that average consensus is achieved. It should be pointed out that, unlike the standard consensus algorithm (4), the tools in nonnegative matrix theory cannot be used to analyze the spectrum of  $M$  directly due to the existence of negative entries.

*Proof of Theorem 1.* (Necessity) Suppose that  $\mathcal{G}$  is not strongly connected. Then at least one node of  $\mathcal{G}$  is not globally reachable. Let  $\mathcal{V}_g^*$  denote the set of non-globally reachable nodes, and write its cardinality  $\text{card}(\mathcal{V}_g^*) = r$ ,  $r \in [1, n]$ . If  $r = n$ , i.e.  $\mathcal{G}$  does not have a globally reachable node, then  $\mathcal{G}$  has at least two distinct closed strong components [27, Theorem 2.1]. In this case, if the nodes in different components have different initial states, then average consensus cannot be achieved. It is left to consider  $r < n$ . Let  $\mathcal{V}_g := \mathcal{V} - \mathcal{V}_g^*$  denote the set of all globally reachable nodes; thus  $\mathcal{V}_g$  is the unique closed strong component in  $\mathcal{G}$  [27, Theorem 2.1]. Consider an initial condition  $(x(0), 0)$  such that all nodes in  $\mathcal{V}_g$  have the same state  $c \in \mathbb{R}$ , and not all the states of the nodes in  $\mathcal{V}_g^*$  equal  $c$ . Hence  $x_a \neq c$ . But no state or surplus update is possible for the nodes in  $\mathcal{V}_g$  because it is closed, and therefore average consensus cannot be achieved.

(Sufficiency) First, we prove the assertion (8). Since  $M_0$  is block (lower) triangular, its spectrum is  $\sigma(M_0) = \sigma(I - L) \cup \sigma(S)$ . Recall that the matrices  $I - L$  and  $S$  are row and column stochastic, respectively; so their spectral radii satisfy  $\rho(I - L) = \rho(S) = 1$ . Now owing to that  $\mathcal{G}$  is strongly connected,  $I - L$  and  $S$  are both *irreducible*; thus by the Perron-Frobenius Theorem (see, e.g., [28, Chapter 8])  $\rho(I - L)$  (resp.  $\rho(S)$ ) is a simple eigenvalue of  $I - L$  (resp.  $S$ ). This implies (8). Moreover, for  $\lambda_1 = \lambda_2 = 1$ , one derives that the corresponding geometric multiplicity equals two by verifying  $\text{rank}(M_0 - I) = 2n - 2$ . Hence the eigenvalue 1 is semi-simple.

Next, we will qualify the changes of the semi-simple eigenvalue  $\lambda_1 = \lambda_2 = 1$  of  $M_0$  under a small perturbation  $\epsilon F$ . We do this by computing the derivatives  $d\lambda_1(\epsilon)/d\epsilon$  and  $d\lambda_2(\epsilon)/d\epsilon$  using Lemma 1; here

$\lambda_1(\epsilon)$  and  $\lambda_2(\epsilon)$  are the eigenvalues of  $M$  corresponding respectively to  $\lambda_1$  and  $\lambda_2$ . To that end, choose the right eigenvectors  $y_1, y_2$  and left eigenvectors  $z_1, z_2$  of the semi-simple eigenvalue 1 as follows:

$$Y := \begin{bmatrix} y_1 & y_2 \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{1} \\ v_2 & -nv_2 \end{bmatrix}, \quad Z := \begin{bmatrix} z_1^T \\ z_2^T \end{bmatrix} = \begin{bmatrix} \mathbf{1}^T & \mathbf{1}^T \\ v_1^T & 0 \end{bmatrix}.$$

Here  $v_1 \in \mathbb{R}^n$  is a left eigenvector of  $I - L$  with respect to  $\rho(I - L)$  such that it is positive and scaled to satisfy  $v_1^T \mathbf{1} = 1$ ; and  $v_2 \in \mathbb{R}^n$  is a right eigenvector of  $S$  corresponding to  $\rho(S)$  such that it is positive and scaled to satisfy  $\mathbf{1}^T v_2 = 1$ . The fact that positive eigenvectors  $v_1$  and  $v_2$  exist follows again from the Perron-Frobenius Theorem. With this choice, one readily checks  $ZY = I$ . Now since  $dM/d\epsilon|_{\epsilon=0} = F$ , the matrix (6) in the present case is

$$\begin{bmatrix} z_1^T F y_1 & z_1^T F y_2 \\ z_2^T F y_1 & z_2^T F y_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ v_1^T v_2 & -nv_1^T v_2 \end{bmatrix}.$$

It follows from Lemma 1 that for small  $\epsilon > 0$ , the derivatives  $d\lambda_1(\epsilon)/d\epsilon$ ,  $d\lambda_2(\epsilon)/d\epsilon$  exist and are the eigenvalues of the above matrix. Hence  $d\lambda_1(\epsilon)/d\epsilon = 0$ , and  $d\lambda_2(\epsilon)/d\epsilon = -nv_1^T v_2 < 0$ . This implies that when  $\epsilon$  is small,  $\lambda_1(\epsilon)$  stays put while  $\lambda_2(\epsilon)$  moves to the left along the real axis. Then by continuity, there must exist a positive  $\delta_1$  such that  $\lambda_1(\delta_1) = 1$  and  $\lambda_2(\delta_1) < 1$ . On the other hand, since eigenvalues are continuous functions of matrix entries (e.g., [21], [22]), there must exist a positive  $\delta_2$  such that  $|\lambda_i(\delta_2)| < 1$  for all  $i \in [3, 2n]$ . Thus for any sufficiently small  $\epsilon \in (0, \min\{\delta_1, \delta_2\})$ , the matrix  $M$  has a simple eigenvalue 1 and all other eigenvalues have moduli smaller than one. Therefore, from Proposition 2, the conclusion that average consensus is achieved follows. ■

*Remark 1.* Assuming that the deterministic algorithm (3) converges to the average, the speed of its convergence is governed by the second largest (in modulus) eigenvalue of the matrix  $M$ . We denote this particular eigenvalue by  $\lambda_2^{(d)}$ , and refer to it as the *convergence factor* of algorithm (3). Note that  $\lambda_2^{(d)}$  depends not only on the graph topology but also on the parameter  $\epsilon$ , and  $\lambda_2^{(d)} < 1$  is equivalent to average consensus (by Proposition 2).

*Remark 2.* Because of adding surpluses, the matrix  $M$  in (3) is double in size and is not nonnegative. Hence standard nonnegative matrix tools cannot be directly applied; this point was also discussed in [12]. In [19] a system matrix containing negative entries was analyzed, which depends however on symmetry of network structures. By contrast, we deal with general network topologies and have demonstrated that certain matrix perturbation tools are useful in proving convergence.

#### D. Proof of Proposition 1

Some preliminaries will be presented first, based on which Proposition 1 follows immediately. Henceforth in this subsection, the digraph  $\mathcal{G}$  is assumed to be strongly connected. We begin by introducing a metric for the distance between the spectrums of  $M_0$  and  $M$ ; here  $M = M_0 + \epsilon F$ , with  $M_0$  and  $F$  in (7). Let  $\sigma(M_0) := \{\lambda_1, \dots, \lambda_{2n}\}$  (where the numbering is the same as that in (8)) and  $\sigma(M) := \{\lambda_1(\epsilon), \dots, \lambda_{2n}(\epsilon)\}$ . The *optimal matching distance*  $d(\sigma(M_0), \sigma(M))$  [21], [22] is defined by

$$d(\sigma(M_0), \sigma(M)) := \min_{\pi} \max_{i \in [1, 2n]} |\lambda_i - \lambda_{\pi(i)}(\epsilon)|, \quad (9)$$

where  $\pi$  is taken over all permutations of  $\{1, \dots, 2n\}$ . Thus if we draw  $2n$  identical circles centered respectively at  $\lambda_1, \dots, \lambda_{2n}$ , then  $d(\sigma(M_0), \sigma(M))$  is the smallest radius such that these circles include all  $\lambda_1(\epsilon), \dots, \lambda_{2n}(\epsilon)$ . Here is an upper bound on the optimal matching distance [21, Theorem VIII.1.5].

*Lemma 2.*  $d(\sigma(M_0), \sigma(M)) \leq 4 (\|M_0\|_{\infty} + \|M\|_{\infty})^{1-1/n} \|\epsilon F\|_{\infty}^{1/n}$ .

Next, we are concerned with the eigenvalues  $\lambda_3(\epsilon), \dots, \lambda_{2n}(\epsilon)$  of  $M$ , whose corresponding unperturbed counterparts  $\lambda_3, \dots, \lambda_{2n}$  of  $M_0$  lie strictly inside the unit circle (see the proof of Theorem 1).

*Lemma 3.* If the parameter  $\epsilon \in (0, \bar{\epsilon}^{(d)})$  with  $\bar{\epsilon}^{(d)}$  in (5), then  $|\lambda_3(\epsilon)|, \dots, |\lambda_{2n}(\epsilon)| < 1$ .

*Proof.* Since  $L = D - A$  and  $S = (I - \tilde{D}) + B$ , one can compute  $\|L\|_{\infty} = 2 \max_{i \in [1, n]} d_i < 2$  and  $\|S\|_{\infty} < n$ . Then  $\|M_0\|_{\infty} \leq \|L\|_{\infty} + \|S\|_{\infty} < 2 + n$  and  $\|F\|_{\infty} \leq 1$ . By Lemma 2,

$$\begin{aligned} d(\sigma(M_0), \sigma(M)) &\leq 4 (2\|M_0\|_{\infty} + \epsilon\|F\|_{\infty})^{1-1/n} (\epsilon\|F\|_{\infty})^{1/n} \\ &< 4 (4 + 2n + \epsilon)^{1-1/n} \epsilon^{1/n} < 4 (4 + 2n + \epsilon) \epsilon^{1/n} < 1 - |\lambda_3|. \end{aligned}$$

The last inequality is due to  $\epsilon < \bar{\epsilon}^{(d)}$  in (5). Now recall from the proof of Theorem 1 that the unperturbed eigenvalues  $\lambda_3, \dots, \lambda_{2n}$  of  $M_0$  lie strictly inside the unit circle; in particular, (8) holds. Therefore, perturbing the eigenvalues  $\lambda_3, \dots, \lambda_{2n}$  by an amount less than  $\bar{\epsilon}$ , the resulting eigenvalues  $\lambda_3(\epsilon), \dots, \lambda_{2n}(\epsilon)$  will remain inside the unit circle. ■

It is left to consider the eigenvalues  $\lambda_1(\epsilon)$  and  $\lambda_2(\epsilon)$  of  $M$ . Since every column sum of  $M$  equals one for an arbitrary  $\epsilon$ , we obtain that 1 is always an eigenvalue of  $M$ . Hence  $\lambda_1(\epsilon)$  must be equal to 1 for any  $\epsilon$ . On the other hand, for  $\lambda_2(\epsilon)$  the following is true.

*Lemma 4.* If the parameter  $\epsilon \in (0, \bar{\epsilon}^{(d)})$  with  $\bar{\epsilon}^{(d)}$  in (5), then  $|\lambda_2(\epsilon)| < 1$ .

*Proof.* First recall from the proof of Theorem 1 that  $\lambda_2 = 1$  and  $d\lambda_2(\epsilon)/d\epsilon < 0$ ; so for sufficiently small  $\epsilon > 0$ , it holds that  $|\lambda_2(\epsilon)| < 1$ . Now suppose that there exists  $\delta \in (0, \bar{\epsilon}^{(d)})$  such that  $|\lambda_2(\delta)| \geq 1$ . Owing

to the continuity of eigenvalues, it suffices to consider  $|\lambda_2(\delta)| = 1$ . There are three such possibilities, for each of which we derive a contradiction.

Case 1:  $\lambda_2(\delta)$  is a complex number with nonzero imaginary part and  $|\lambda_2(\delta)| = 1$ . Since  $M$  is a real matrix, there must exist another eigenvalue  $\lambda_i(\delta)$ , for some  $i \in [3, 2n]$ , such that  $\lambda_i(\delta)$  is a complex conjugate of  $\lambda_2(\delta)$ . Then  $|\lambda_i(\delta)| = |\lambda_2(\delta)| = 1$ , which is in contradiction to that all the eigenvalues  $\lambda_3(\delta), \dots, \lambda_{2n}(\delta)$  stay inside the unit circle as  $\delta \in (0, \bar{\epsilon}^{(d)})$  by Lemma 3.

Case 2:  $\lambda_2(\delta) = -1$ . This implies at least  $d(\sigma(M_0), \sigma(M)) = 2$ , which contradicts  $d(\sigma(M_0), \sigma(M)) < 1 - |\lambda_3| < 1$  when (5) holds.

Case 3:  $\lambda_2(\delta) = 1$ . This case is impossible because the eigenvalue 1 of  $M$  is always simple, as we have justified in the necessity proof of Proposition 2. ■

Summarizing Lemmas 3 and 4, we obtain that if the parameter  $\epsilon \in (0, \bar{\epsilon}^{(d)})$  with  $\bar{\epsilon}^{(d)}$  in (5), then  $\lambda_1(\epsilon) = 1$  and  $|\lambda_2(\epsilon)|, |\lambda_3(\epsilon)|, \dots, |\lambda_{2n}(\epsilon)| < 1$ . Therefore, by Proposition 2 the deterministic algorithm (3) achieves average consensus; this establishes Proposition 1.

#### IV. AVERAGING IN ASYNCHRONOUS NETWORKS

We move on to solve Problem 2. First, a surplus-based gossip algorithm is designed for digraphs, which extends those algorithms [6]–[9] only for undirected graphs. Then, mean-square and almost sure convergence to average consensus is justified for arbitrary strongly connected topologies.

##### A. Algorithm Description

Consider again a network of  $n$  agents modeled by a digraph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ . Suppose that at each time, exactly one edge in  $\mathcal{E}$  is activated at random, independently from all earlier instants. Say edge  $(j, i)$  is activated at time  $k \in \mathbb{Z}_+$ , with a constant probability  $p_{ij} \in (0, 1)$ . Along the edge, the state information  $x_j(k)$  and surplus  $s_j(k)$  are transmitted from node  $j$  to  $i$ . The induced update is described as follows:

(i) Let  $w_{ij} \in (0, 1)$  be the updating weight, and  $\epsilon > 0$  be a parameter. For node  $i$ :

$$x_i(k+1) = x_i(k) + w_{ij}(x_j(k) - x_i(k)) + \epsilon w_{ij} s_i(k), \quad (10)$$

$$s_i(k+1) = s_i(k) + s_j(k) - (x_i(k+1) - x_i(k)), \quad (11)$$

(ii) For node  $j$ :  $x_j(k+1) = x_j(k)$  and  $s_j(k+1) = 0$ .

(iii) For other nodes  $l \in \mathcal{V} - \{i, j\}$ :  $x_l(k+1) = x_l(k)$  and  $s_l(k+1) = s_l(k)$ .

We discuss potential applications of this protocol in sensor networks. Our focus is again on the situations of directed information flow, like asynchronous communication with variable ranges or unidirectional

geographic routing [13], [14]. First, the states and surpluses can be implemented as ordinary variables in sensors, since their exchange and updating rules are fairly simple and purely local. Also, we will see that the parameter  $\epsilon$ , as in the algorithm (3), affects the convergence of the algorithm, and must be chosen to be sufficiently small. A valid upper bound for  $\epsilon$  involves again non-local information of the network; thus computing a bound offline and then notifying that value to every node is one possible way to deal with this restriction.

Now let  $A_{ji}$  be the adjacency matrix of the digraph  $\mathcal{G}_{ji} = (\mathcal{V}, \{(j, i)\})$  given by  $A_{ji} = w_{ij} f_i f_j^T$ , where  $f_i, f_j$  are unit vectors of the standard basis of  $\mathbb{R}^n$ . Then the Laplacian matrix  $L_{ji}$  is given by  $L_{ji} := D_{ji} - A_{ji}$ , where  $D_{ji} = w_{ij} f_i f_i^T$ . Thus  $L_{ji}$  has zero row sums, and the matrix  $I - L_{ji}$  is row stochastic. Also define  $S_{ji} := I - (f_j - f_i) f_j^T$ ; it is clear that  $S_{ji}$  is column stochastic. With these matrices, the iteration of states and surpluses when edge  $(j, i)$  is activated at time  $k$  can be written in the matrix form as

$$\begin{bmatrix} x(k+1) \\ s(k+1) \end{bmatrix} = M(k) \begin{bmatrix} x(k) \\ s(k) \end{bmatrix}, \text{ where } M(k) = M_{ji} := \begin{bmatrix} I - L_{ji} & \epsilon D_{ji} \\ L_{ji} & S_{ji} - \epsilon D_{ji} \end{bmatrix}. \quad (12)$$

We have several remarks regarding this algorithm. (i) The matrix  $M(k)$  has negative entries due to the presence of the Laplacian matrix  $L_{ji}$  in the  $(2, 1)$ -block. (ii) The column sums of  $M(k)$  are equal to one, which implies that the quantity  $x(k) + s(k)$  is constant for all  $k$ . (iii) By the assumption on the probability distribution of activating edges, the sequence  $M(k)$ ,  $k = 0, 1, \dots$ , is independent and identically distributed (i.i.d.). Henceforth we refer to (12) as the gossip algorithm, and establish its mean-square and almost sure convergence in the sequel.

*Example 2.* Consider again the network of four nodes in Fig. 1. We give one instance of the matrix  $M(k)$  when edges  $(3, 2)$  is activated, with the updating weight  $w_{23} = 1/2$ .

$$M_{32} = \left[ \begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 & \epsilon/2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1/2 & -1/2 & 0 & 0 & 1 - \epsilon/2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right],$$

We see that  $M(k)$  has negative entries, and every column sums up to one.

### B. Convergence Result

We present our main result in this section.

*Theorem 2.* Using the gossip algorithm (12) with the parameter  $\epsilon > 0$  sufficiently small, the agents achieve mean-square average consensus if and only if the digraph  $\mathcal{G}$  is strongly connected.

We remark that Theorem 2 generalizes the convergence result in [6] from undirected to directed graphs. The problem of achieving average consensus on gossip digraphs is, however, more difficult in that the state sum of the nodes need not be invariant at each iteration. The key in our extension is to augment surplus variables which keep track of individual state updates, thereby ensuring average consensus for general strongly connected digraphs. This approach was previously exploited in [12] for a broadcast gossip algorithm, however without a convergence proof. We remark that our technique to prove Theorem 2, based on matrix perturbation theory, can be applied to [12] and justify the algorithm convergence.

We note that in the literature, many works for agents with non-scalar dynamics deal only with static networks (e.g., [25], [29]). Some exceptions include [19] which relies heavily on graph symmetry and [5] which is based on dwell-time switching. By contrast, we study general digraphs that switch at every discrete time and each resulting update matrix is not nonnegative. The corresponding analysis is difficult, and we will demonstrate again that matrix perturbation tools are instrumental in proving convergence.

To prove Theorem 2, three preliminary results are to be established in order. The first is a necessary and sufficient condition for mean-square average consensus characterized by the spectrum of the matrix  $E[M(k) \otimes M(k)]$ , where  $\otimes$  stands for the Kronecker product. This condition will be used in the sufficiency proof of Theorem 2. Since the matrices  $M(k)$  are i.i.d. we denote  $E[M(k) \otimes M(k)]$  by  $E[M \otimes M]$ . This result corresponds to Proposition 2 for the deterministic algorithm in Section III.

*Proposition 3.* The gossip algorithm (12) achieves mean-square average consensus if and only if 1 is a simple eigenvalue of  $E[M \otimes M]$ , and all the other eigenvalues have moduli smaller than one.

*Proof.* (Sufficiency) Define the consensus error  $e(k)$ ,  $k \geq 0$ , as

$$e(k) := \begin{bmatrix} x(k) \\ s(k) \end{bmatrix} - \begin{bmatrix} x_a \mathbf{1} \\ 0 \end{bmatrix} \in \mathbb{R}^{2n}. \quad (13)$$

We must show that  $E[e(k)^T e(k)] \rightarrow 0$  as  $k \rightarrow \infty$ . Since  $\mathbf{1}^T(x(k) + s(k)) = \mathbf{1}^T x(0)$  for every  $k \geq 0$ ,  $e(k)$  is orthogonal to  $[\mathbf{1}^T \ \mathbf{1}^T]^T$  (i.e.,  $[\mathbf{1}^T \ \mathbf{1}^T]e(k) = 0$ ). Also it is easy to check  $e(k+1) = M(k)e(k)$ ; thus  $e(k+1)e(k+1)^T = M(k)e(k)e(k)^T M(k)^T$ . Collect the entries of  $e(k)e(k)^T$ , drawn column wise, into a vector  $\tilde{e}(k) \in \mathbb{R}^{4n^2}$ . It then suffices to show that  $E[\tilde{e}(k)] \rightarrow 0$  as  $k \rightarrow \infty$ .

Now it follows that  $\tilde{e}(k+1) = (M(k) \otimes M(k)) \tilde{e}(k)$  (cf. [6]). Hence  $E[\tilde{e}(k+1)|\tilde{e}(k)] = E[M \otimes M] \tilde{e}(k)$ , and condition repeatedly to obtain  $E[\tilde{e}(k)] = E[M \otimes M]^k \tilde{e}(0)$ . Note that every column of  $E[M \otimes M]$  sums up to one, and

$$E[M \otimes M] \left( \begin{bmatrix} \mathbf{1} \\ 0 \end{bmatrix} \otimes \begin{bmatrix} \mathbf{1} \\ 0 \end{bmatrix} \right) = \begin{bmatrix} \mathbf{1} \\ 0 \end{bmatrix} \otimes \begin{bmatrix} \mathbf{1} \\ 0 \end{bmatrix};$$

so 1 is an eigenvalue of  $E[M \otimes M]$ , with  $[\mathbf{1}^T \ \mathbf{1}^T]^T \otimes [\mathbf{1}^T \ \mathbf{1}^T]^T$  and  $[\mathbf{1}^T \ 0]^T \otimes [\mathbf{1}^T \ 0]^T$  as associated left and right eigenvectors, respectively. Write  $E[M \otimes M]$  in Jordan canonical form as

$$E[M \otimes M] = VJV^{-1} = \begin{bmatrix} y_1 & \cdots & y_{4n^2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & J' \end{bmatrix} \begin{bmatrix} z_1^T \\ \vdots \\ z_{4n^2}^T \end{bmatrix},$$

where  $J'$  contains the Jordan block matrices corresponding to those eigenvalues with moduli smaller than one. For the eigenvalue 1 choose  $y_1 = [\mathbf{1}^T \ 0]^T \otimes [\mathbf{1}^T \ 0]^T$  and  $z_1 = 1/n^2 [\mathbf{1}^T \ \mathbf{1}^T]^T \otimes [\mathbf{1}^T \ \mathbf{1}^T]^T$ ; thus  $z_1^T y_1 = 1$ . Then the  $k$ th power of  $E[M \otimes M]$  is

$$E[M \otimes M]^k = VJ^kV^{-1} = V \begin{bmatrix} 1 & 0 \\ 0 & (J')^k \end{bmatrix} V^{-1} \rightarrow y_1 z_1^T, \quad \text{as } k \rightarrow \infty.$$

Therefore we obtain

$$E[\tilde{e}(k)] \rightarrow y_1 z_1^T \tilde{e}(0) = y_1 \sum_{i=1}^{2n} \left( e_i(0) \sum_{j=1}^{2n} e_j(0) \right) = y_1 \sum_{i=1}^{2n} e_i(0) \cdot 0 = 0,$$

where the second equality is due to  $e(k) \perp [\mathbf{1}^T \ \mathbf{1}^T]^T$ .

(Necessity) Suppose  $E[e(k)^T e(k)] \rightarrow 0$  as  $k \rightarrow \infty$ . Then  $E[e_i(k)^2] \rightarrow 0$  for all  $i$ . It thus follows from the Cauchy-Schwartz inequality (e.g., [24]) that  $E[|e_i(k)e_j(k)|]^2 \leq E[e_i(k)^2] E[e_j(k)^2] \rightarrow 0$ , for every  $i, j \in [1, 2n]$ . This implies  $E[\tilde{e}(k)] \rightarrow 0$ ; so  $\lim_{k \rightarrow \infty} E[M \otimes M]^k \tilde{e}(0) = 0$ . Also, it is known [4] that  $\lim_{k \rightarrow \infty} E[M \otimes M]^k$  exists if and only if there is a nonsingular  $V$  such that

$$E[M \otimes M] = VJV^{-1} = \begin{bmatrix} y_1 & \cdots & y_{4n^2} \end{bmatrix} \begin{bmatrix} I_\kappa & 0 \\ 0 & J' \end{bmatrix} \begin{bmatrix} z_1^T \\ \vdots \\ z_{4n^2}^T \end{bmatrix},$$

where  $\kappa \in [1, 2n]$  and  $\rho(J') < 1$ . Hence  $\lim_{k \rightarrow \infty} E[M \otimes M]^k \tilde{e}(0) = (\sum_{i=1}^{\kappa} y_i z_i^T) \tilde{e}(0) = 0$ . Now suppose  $\kappa > 1$ . Choose as before  $z_1 = 1/n^2 [\mathbf{1}^T \ \mathbf{1}^T]^T \otimes [\mathbf{1}^T \ \mathbf{1}^T]^T$ , and recall  $z_1^T e(0) = 0$ . We know from the structure of  $J$  that for every  $j \in [2, \kappa]$ ,  $z_j$  is linearly independent of  $z_1$ , which indicates  $z_j^T e(0) \neq 0$  and consequently  $(\sum_{i=1}^{\kappa} y_i z_i^T) \tilde{e}(0) \neq 0$ . Therefore  $\kappa = 1$ , i.e., the eigenvalue 1 of  $E[M \otimes M]$  is simple and all the others have moduli smaller than one. ■

The second preliminary is an easy corollary of the Perron-Frobenius Theorem.

*Lemma 5.* (cf. [30, Chapter XIII]) Let  $W$  be a nonnegative and irreducible matrix, and  $\lambda$  be an eigenvalue of  $W$ . If there is a positive vector  $v$  such that  $Wv = \lambda v$ , then  $\lambda = \rho(W)$ .

*Proof.* Since  $W$  is nonnegative and irreducible, the Perron-Frobenius Theorem implies that  $\rho(W)$  is a simple eigenvalue of  $W$  and there is a positive left eigenvector  $w$  corresponding to  $\rho(W)$ , i.e.,  $w^T W = w^T \rho(W)$ . Then

$$\rho(W)(v^T w) = v^T (\rho(W)w) = v^T (W^T w) = (Wv)^T w = (\lambda v)^T w = \lambda(v^T w),$$

which yields  $(\lambda - \rho(W))(v^T w) = 0$ . Since both  $v$  and  $w$  are positive, we conclude that  $\lambda = \rho(W)$ . ■

The last preliminary is on the spectral properties of the following four matrices:  $E[(I - L) \otimes (I - L)]$ ,  $E[(I - L) \otimes S]$ ,  $E[S \otimes (I - L)]$ , and  $E[S \otimes S]$ .

*Lemma 6.* Suppose that the digraph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  is strongly connected. Then each of the four matrices  $E[(I - L) \otimes (I - L)]$ ,  $E[(I - L) \otimes S]$ ,  $E[S \otimes (I - L)]$ , and  $E[S \otimes S]$  has a simple eigenvalue 1 and all other eigenvalues with moduli smaller than one.

*Proof.* First observe that all the four matrices are nonnegative, for  $I - L_{ji}$  and  $S_{ji}$  are for every  $(j, i) \in \mathcal{E}$ . Then since  $(I - L_{ji})\mathbf{1} = \mathbf{1}$  and  $\mathbf{1}^T S_{ji} = \mathbf{1}^T$  for every  $(j, i) \in \mathcal{E}$ , a short calculation yields the following:

$$\begin{aligned} E[(I - L) \otimes (I - L)](\mathbf{1} \otimes \mathbf{1}) &= (\mathbf{1} \otimes \mathbf{1}); & E[(I - L) \otimes S](\mathbf{1} \otimes v_2) &= (\mathbf{1} \otimes v_2); \\ (\mathbf{1}^T \otimes v_1^T)E[S \otimes (I - L)] &= (\mathbf{1}^T \otimes v_1^T); & (\mathbf{1}^T \otimes \mathbf{1}^T)E[S \otimes S] &= (\mathbf{1}^T \otimes \mathbf{1}^T). \end{aligned}$$

Here  $v_1$  is positive such that  $v_1^T E[I - L] = v_1^T$  and  $v_1^T \mathbf{1} = 1$ , and  $v_2$  is positive such that  $E[S]v_2 = v_2$  and  $\mathbf{1}^T v_2 = 1$ . Thus each matrix has an eigenvalue 1, and the corresponding right or left eigenvector is positive. In what follows, it will be shown that all the four matrices are irreducible. Then the conclusion will follow from Lemma 5 and the Perron-Frobenius Theorem.

We first prove that  $E[(I - L) \otimes (I - L)]$  is irreducible, which is equivalent to that the digraph  $\hat{\mathcal{G}} = (\hat{\mathcal{V}}, \hat{\mathcal{E}})$  corresponding to this matrix is strongly connected, where  $\hat{\mathcal{V}} := \mathcal{V} \times \mathcal{V} = \{(i, i') : i, i' \in \mathcal{V}\}$ . Arrange the nodes in  $\hat{\mathcal{V}}$  so that  $\hat{\mathcal{V}} = \mathcal{V}_1 \cup \dots \cup \mathcal{V}_n$ , where  $\mathcal{V}_p = \{(p, 1), \dots, (p, n)\}$  for every  $p \in [1, n]$ . Now since  $E[(I - L) \otimes (I - L)] = \sum_{(j, i) \in \mathcal{E}} p_{ij}(I - L_{ji}) \otimes (I - L_{ji})$ , the digraph  $\hat{\mathcal{G}}$  is the union of the digraphs corresponding to  $p_{ij}(I - L_{ji}) \otimes (I - L_{ji})$ . Note that each  $p_{ij}(I - L_{ji}) \otimes (I - L_{ji})$  gives rise to (i) an edge from  $(p, j)$  to  $(p, i)$  in  $\mathcal{V}_p$  for every  $p \in [1, n]$ , and (ii) edges from some nodes in  $\mathcal{V}_j$  to some nodes in  $\mathcal{V}_i$ . Owing to that  $\mathcal{G}$  is strongly connected, the union of the above edges yields, for every  $i, j \in [1, n]$ , (i) a directed path from  $(p, i)$  to  $(p, j)$  in  $\mathcal{V}_p$  for every  $p \in [1, n]$ , and (ii) directed paths from

some nodes in  $\mathcal{V}_i$  to some nodes in  $\mathcal{V}_j$ . This implies that there is a directed path from  $(p, i)$  to  $(q, j)$  for every  $p, q, i, j \in [1, n]$ , i.e.,  $\hat{\mathcal{G}}$  is strongly connected, and hence  $E[(I - L) \otimes (I - L)]$  is irreducible.

By a similar argument, we derive that the digraphs corresponding to  $E[(I - L) \otimes S]$ ,  $E[S \otimes (I - L)]$ , and  $E[S \otimes S]$  are all strongly connected. Therefore they are also irreducible. ■

We are now ready to provide the proof of Theorem 2. The necessity argument is the same as Theorem 1. Below is the sufficiency part.

*Proof of Theorem 2. (Sufficiency)* By Proposition 3 it suffices to show that the matrix  $E[M \otimes M]$  has a simple eigenvalue 1, and all other eigenvalues with moduli smaller than one. Let  $M_0(k) := \begin{bmatrix} I - L(k) & 0 \\ L(k) & S(k) \end{bmatrix}$  and  $F(k) := \begin{bmatrix} 0 & D(k) \\ 0 & -D(k) \end{bmatrix}$ ; from (12) we have  $M(k) = M_0(k) + \epsilon F(k)$ . Then write

$$\begin{aligned} E[M \otimes M] &= E[(M_0 + \epsilon F) \otimes (M_0 + \epsilon F)] = E[M_0 \otimes M_0] + \epsilon E[M_0 \otimes F + F \otimes M_0 + F \otimes \epsilon F] \\ &= E\left\{ \begin{bmatrix} I - L & 0 \\ L & S \end{bmatrix} \otimes \begin{bmatrix} I - L & 0 \\ L & S \end{bmatrix} \right\} + \epsilon E\left\{ \begin{bmatrix} I - L & 0 \\ L & S \end{bmatrix} \otimes \begin{bmatrix} 0 & D \\ 0 & -D \end{bmatrix} + \begin{bmatrix} 0 & D \\ 0 & -D \end{bmatrix} \otimes \begin{bmatrix} I - L & 0 \\ L & S \end{bmatrix} + \begin{bmatrix} 0 & D \\ 0 & -D \end{bmatrix} \otimes \epsilon \begin{bmatrix} 0 & D \\ 0 & -D \end{bmatrix} \right\}. \end{aligned}$$

Let  $p \in [1, 4n]$ , and  $p\mathbf{n} := \{(p-1)n+1, \dots, pn\}$ . Consider the following permutation:

$$\begin{aligned} &\{\mathbf{n}, 3\mathbf{n}, \dots, (2n-1)\mathbf{n}; 2\mathbf{n}, 4\mathbf{n}, \dots, 2n\mathbf{n}; \\ &(2n+1)\mathbf{n}, (2n+3)\mathbf{n}, \dots, (4n-1)\mathbf{n}; (2n+2)\mathbf{n}, (2n+4)\mathbf{n}, \dots, 4n\mathbf{n}\}. \end{aligned}$$

Denoting by  $P$  the corresponding permutation matrix (which is orthogonal), one derives that

$$P^T E[M \otimes M] P = P^T E[M_0 \otimes M_0] P + \epsilon P^T E[M_0 \otimes F + F \otimes M_0 + F \otimes \epsilon F] P =: \hat{M}_0 + \epsilon \hat{F}, \quad (14)$$

where

$$\begin{aligned} \hat{M}_0 &:= E \begin{bmatrix} (I - L) \otimes (I - L) & 0 & 0 & 0 \\ (I - L) \otimes L & (I - L) \otimes S & 0 & 0 \\ L \otimes (I - L) & 0 & S \otimes (I - L) & 0 \\ L \otimes L & L \otimes S & S \otimes L & S \otimes S \end{bmatrix}, \\ \hat{F} &:= E \begin{bmatrix} 0 & (I - L) \otimes D & D \otimes (I - L) & D \otimes \epsilon D \\ 0 & -(I - L) \otimes D & D \otimes L & D \otimes (S - \epsilon D) \\ 0 & L \otimes D & -D \otimes (I - L) & (S - \epsilon D) \otimes D \\ 0 & -L \otimes D & -D \otimes L & D \otimes (\epsilon D - S) - S \otimes D \end{bmatrix}. \end{aligned}$$

Based on the above similarity transformation, we henceforth analyze the spectral properties of the matrix  $\hat{M}_0 + \epsilon \hat{F}$ . For this, we resort again to a perturbation argument, which proceeds similarly to the one for Theorem 1. First, since the digraph  $\mathcal{G}$  is strongly connected, it follows from Lemma 6 that the eigenvalues of the matrix  $\hat{M}_0$  satisfy

$$1 = \hat{\lambda}_1 = \hat{\lambda}_2 = \hat{\lambda}_3 = \hat{\lambda}_4 > |\hat{\lambda}_5| \geq \dots \geq |\hat{\lambda}_{4n^2}|. \quad (15)$$

For the eigenvalue 1, one derives that the corresponding geometric multiplicity equals four by verifying  $\text{rank}(\hat{M}_0 - I) = 4n^2 - 4$ . Thus 1 is a semi-simple eigenvalue.

Next, we will qualify the changes of the semi-simple eigenvalue  $\hat{\lambda}_1 = \hat{\lambda}_2 = \hat{\lambda}_3 = \hat{\lambda}_4 = 1$  of  $\hat{M}_0$  under a small perturbation  $\epsilon \hat{F}$ . We do this by computing the derivatives  $d\hat{\lambda}_i(\epsilon)/d\epsilon$ ,  $i \in [1, 4]$ , using Lemma 1; here  $\hat{\lambda}_i(\epsilon)$  are the eigenvalues of  $\hat{M}_0 + \epsilon \hat{F}$  corresponding to  $\hat{\lambda}_i$ . To that end, choose the right and left eigenvectors of the semi-simple eigenvalue 1 as follows:

$$Y := \begin{bmatrix} y_1 & y_2 & y_3 & y_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & \mathbf{1} \otimes \mathbf{1} \\ 0 & 0 & \mathbf{1} \otimes nv_2 & -\mathbf{1} \otimes nv_2 \\ 0 & nv_2 \otimes \mathbf{1} & 0 & -nv_2 \otimes \mathbf{1} \\ nv_2 \otimes nv_2 & -nv_2 \otimes nv_2 & -nv_2 \otimes nv_2 & nv_2 \otimes nv_2 \end{bmatrix},$$

$$Z := \begin{bmatrix} z_1^T \\ z_2^T \\ z_3^T \\ z_4^T \end{bmatrix} = \begin{bmatrix} \frac{1}{n} \mathbf{1}^T \otimes \frac{1}{n} \mathbf{1}^T & \frac{1}{n} \mathbf{1}^T \otimes \frac{1}{n} \mathbf{1}^T & \frac{1}{n} \mathbf{1}^T \otimes \frac{1}{n} \mathbf{1}^T & \frac{1}{n} \mathbf{1}^T \otimes \frac{1}{n} \mathbf{1}^T \\ \frac{1}{n} \mathbf{1}^T \otimes v_1^T & 0 & \frac{1}{n} \mathbf{1}^T \otimes v_1^T & 0 \\ v_1^T \otimes \frac{1}{n} \mathbf{1}^T & v_1^T \otimes \frac{1}{n} \mathbf{1}^T & 0 & 0 \\ v_1^T \otimes v_1^T & 0 & 0 & 0 \end{bmatrix}.$$

Here  $v_1$  is positive such that  $v_1^T E[I - L] = v_1^T$  and  $v_1^T \mathbf{1} = 1$ , and  $v_2$  is positive such that  $E[S] v_2 = v_2$  and  $\mathbf{1}^T v_2 = 1$ . With this choice, it is readily checked that  $ZY = I$ . Now the matrix  $\hat{M}_0 + \epsilon \hat{F}$  depends smoothly on  $\epsilon$ , and the derivative  $d(\hat{M}_0 + \epsilon \hat{F})/d\epsilon|_{\epsilon=0}$  is

$$\hat{F}_0 := \left. \frac{d(\hat{M}_0 + \epsilon \hat{F})}{d\epsilon} \right|_{\epsilon=0} = \left( \hat{F} + \epsilon \frac{d\hat{F}}{d\epsilon} \right) \Big|_{\epsilon=0}$$

$$= E \begin{bmatrix} 0 & (I - L) \otimes D & D \otimes (I - L) & 0 \\ 0 & -(I - L) \otimes D & D \otimes L & D \otimes S \\ 0 & L \otimes D & -D \otimes (I - L) & S \otimes D \\ 0 & -L \otimes D & -D \otimes L & -D \otimes S - S \otimes D \end{bmatrix}.$$

Hence the matrix (6) in the present case is

$$\begin{bmatrix} z_1^T \hat{F}_0 y_1 & z_1^T \hat{F}_0 y_2 & z_1^T \hat{F}_0 y_3 & z_1^T \hat{F}_0 y_4 \\ z_2^T \hat{F}_0 y_1 & z_2^T \hat{F}_0 y_2 & z_2^T \hat{F}_0 y_3 & z_2^T \hat{F}_0 y_4 \\ z_3^T \hat{F}_0 y_1 & z_3^T \hat{F}_0 y_2 & z_3^T \hat{F}_0 y_3 & z_3^T \hat{F}_0 y_4 \\ z_4^T \hat{F}_0 y_1 & z_4^T \hat{F}_0 y_2 & z_4^T \hat{F}_0 y_3 & z_4^T \hat{F}_0 y_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ nv_1^T E[D]v_2 & -nv_1^T E[D]v_2 & 0 & 0 \\ nv_1^T E[D]v_2 & 0 & -nv_1^T E[D]v_2 & 0 \\ 0 & nv_1^T E[D]v_2 & nv_1^T E[D]v_2 & -2nv_1^T E[D]v_2 \end{bmatrix}.$$

It follows from Lemma 1 that for small  $\epsilon > 0$ , the derivatives  $d\hat{\lambda}_i(\epsilon)/d\epsilon$ ,  $i \in [1, 4]$ , exist and are the eigenvalues of the above matrix. Hence  $d\hat{\lambda}_1(\epsilon)/d\epsilon = 0$ ,  $d\hat{\lambda}_2(\epsilon)/d\epsilon = d\hat{\lambda}_3(\epsilon)/d\epsilon = -nv_1^T E[D]v_2 < 0$ , and  $d\hat{\lambda}_4(\epsilon)/d\epsilon = -2nv_1^T E[D]v_2 < 0$ . This implies that when  $\epsilon$  is small,  $\hat{\lambda}_1(\epsilon)$  stays put, while  $\hat{\lambda}_2(\epsilon)$ ,  $\hat{\lambda}_3(\epsilon)$ , and  $\hat{\lambda}_4(\epsilon)$  move to the left along the real axis. So by continuity, there exists a positive  $\delta_1$  such that  $\lambda_1(\delta_1) = 1$  and  $\lambda_2(\delta_1), \lambda_3(\delta_1), \lambda_4(\delta_1) < 1$ . On the other hand, by the eigenvalue continuity there exists a positive  $\delta_2$  such that  $|\lambda_i(\delta_2)| < 1$  for all  $i \in [5, 4n^2]$ . Therefore for any sufficiently small  $\epsilon \in (0, \min\{\delta_1, \delta_2\})$ , the matrix  $\hat{M}_0 + \epsilon \hat{F}$  has a simple eigenvalue 1 and all other eigenvalues with moduli smaller than one. ■

*Remark 3.* Assuming that the gossip algorithm (12) converges to the average in mean square, the speed of its convergence is determined by the second largest (in modulus) eigenvalue of the matrix  $E[M \otimes M]$ . We denote this particular eigenvalue by  $\lambda_2^{(g)}$ , and refer to it as the *convergence factor* of algorithm (12). Notice that  $\lambda_2^{(g)}$  depends not only on the graph topology but also on the parameter  $\epsilon$ , and  $\lambda_2^{(g)} < 1$  is equivalent to mean-square average consensus (by Proposition 3).

*Remark 4.* We have established that for small enough  $\epsilon$ , the gossip algorithm (12) achieves mean-square average consensus. Using the same notion of optimal matching distance and following the procedures as in Subsection III-D, it may be possible to derive a general bound for  $\epsilon$  by solving the inequality  $4(\|\hat{M}_0\|_\infty + \|\hat{M}_0 + \epsilon \hat{F}\|_\infty)^{1-1/n} \|\epsilon \hat{F}\|_\infty^{1/n} < 1 - |\hat{\lambda}_5|$ , where  $\hat{M}_0, \hat{F}$  are from (14) and  $\hat{\lambda}_5$  from (15). The corresponding computation is however rather long, since the involved matrices are of much larger sizes. Such a general bound unavoidably again involves  $n$ , the number of agents in the network, and  $\hat{\lambda}_5$ , the second largest eigenvalue of one of the four matrices in Lemma 6. Consequently, the bound for  $\epsilon$  is conservative and requires the structure of the network.

Finally, we consider almost sure average consensus. Note that the gossip algorithm (12) can be viewed as a *jump linear system*, with i.i.d. system matrices  $M(k)$ ,  $k \in \mathbb{Z}_+$ . For such systems, it is known (e.g., [31, Corollary 3.46]) that almost sure convergence can be implied from mean-square convergence. Therefore the result on almost sure convergence is immediate.

*Corollary 1.* Using the gossip algorithm (12) with the parameter  $\epsilon > 0$  sufficiently small, the agents achieve almost sure average consensus if and only if the digraph  $\mathcal{G}$  is strongly connected.

## V. SPECIAL TOPOLOGIES

We turn now to a special class of topologies – strongly connected and balanced digraphs – and investigate the required upper bound on the parameter  $\epsilon$  for the deterministic algorithm (3). Furthermore, when these digraphs are restricted to symmetric or cyclic respectively, we derive less conservative  $\epsilon$  bounds compared to the general one in (5).

Given a digraph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , its *degree*  $d$  is defined by  $d := \max_{i \in \mathcal{V}} \text{card}(\mathcal{N}_i^+)$ . In the deterministic algorithm (3) choose the updating and sending weights to be respectively  $a_{ij} = 1/(2dn)$  and  $b_{ij} = 1/(dn)$ , for every  $(j, i) \in \mathcal{E}$ . This choice renders the two matrices  $I - 2L$  and  $S$  identical, when the digraph  $\mathcal{G}$  is balanced. We will see that the equality  $I - 2L = S$  supports a similarity transformation in dealing with the cyclic case below.

*Lemma 7.* Suppose that the parameter  $\epsilon$  satisfies  $\epsilon \in (0, 2)$ , and the zeros of the following polynomial for every  $\mu \neq 0$  with  $|\mu - 1/(2n)| \leq 1/(2n)$  lie strictly inside the unit circle:

$$p(\lambda) := \lambda^2 + \alpha_1 \lambda + \alpha_0, \quad (16)$$

where  $\alpha_0 := 2\mu^2 - 3\mu - \epsilon + 1$ ,  $\alpha_1 := 3\mu + \epsilon - 2$ . Then the deterministic algorithm (3) achieves average consensus on strongly connected and balanced digraphs.

*Proof.* We analyze the spectral properties of the matrix  $M$  in terms of those of the Laplacian matrix  $L$ . Let  $\mu_i$ ,  $i = 1, \dots, n$ , be the  $i$ th eigenvalue of  $L$ . Since  $\mathcal{G}$  is balanced and all the updating weights are  $a_{ij} = 1/(2dn)$ , it follows from the Gershgorin Theorem [28, Chapter 6] that  $|\mu_i - 1/(2n)| \leq 1/(2n)$ . Further, as  $\mathcal{G}$  is strongly connected, by the Perron-Frobenius Theorem [28, Chapter 8] we get that  $\mu_1 = 0$  is simple. Now substituting the equality  $S = I - 2L$  into (3) one obtains

$$M = \begin{bmatrix} I - L & \epsilon I \\ L & I - 2L - \epsilon I \end{bmatrix}.$$

Consider the characteristic polynomial of  $M$ :

$$\begin{aligned} \det(\lambda I - M) &= \det \left( \begin{bmatrix} (\lambda - 1)I + L & -\epsilon I \\ -L & (\lambda - 1 + \epsilon)I + 2L \end{bmatrix} \right) \\ &= \det(((\lambda - 1)I + L)((\lambda - 1 + \epsilon)I + 2L) - \epsilon L) \\ &= \det((\lambda - 1)(\lambda - 1 + \epsilon)I + 3(\lambda - 1)L + 2L^2). \end{aligned}$$

Here the second equality is due to that  $(\lambda - 1)I + L$  and  $-L$  commute [32]. By spectral mapping one derives that the  $2n$  eigenvalues of  $M$  can be obtained by solving the following  $n$  equations:

$$(\lambda - 1)(\lambda - 1 + \epsilon) + 3(\lambda - 1)\mu_i + 2\mu_i^2 = 0, \quad i = 1, \dots, n. \quad (17)$$

For  $\mu_1 = 0$  we have from (17) that  $\lambda_1 = 1$  and  $\lambda_2 = 1 - \epsilon$ . Since  $\epsilon \in (0, 2)$ ,  $\lambda_2 \in (-1, 1)$ . Now fix  $i \in [2, n]$  so that  $\mu_i \neq 0$  and  $|\mu_i - 1/(2n)| \leq 1/(2n)$ . Note that the left hand side of (17) can be arranged into the polynomial  $p(\lambda)$  in (16), whose zeros are inside the unit circle. It follows that 1 is a simple eigenvalue of  $M$ , and all other eigenvalues have moduli smaller than one. Therefore, by Proposition 2 we conclude that average consensus is achieved. ■

Now we investigate the values of  $\epsilon$  that ensure the zeros of the polynomial  $p(\lambda)$  in (16) inside the unit circle, which in turn guarantee average consensus on strongly connected and balanced digraphs by Lemma 7. For this, we view the polynomial  $p(\lambda)$  as *interval polynomials* [33] by letting  $\mu$  take any value in the square:  $0 \leq \operatorname{Re}(\mu) \leq 1/n$ ,  $-1/(2n) \leq \operatorname{Im}(\mu) \leq 1/(2n)$ . Applying the bilinear transformation we obtain a new family of interval polynomials:

$$\tilde{p}(\gamma) := (\gamma - 1)^2 p\left(\frac{\gamma + 1}{\gamma - 1}\right) = (1 + \alpha_0 + \alpha_1)\gamma^2 + (2 - 2\alpha_0)\gamma + (1 + \alpha_0 - \alpha_1). \quad (18)$$

Then by Kharitonov's result for the complex-coefficient case, the stability of  $\tilde{p}(\gamma)$  (its zeros have negative real parts) is equivalent to the stability of eight extreme polynomials [33, Section 6.9], which in turn suffices to guarantee that the zeros of  $p(\lambda)$  lie strictly inside the unit circle. Checking the stability of eight extreme polynomials results in upper bounds on  $\epsilon$  in terms of  $n$ . This is displayed in Fig. 2 as the solid curve. We see that the bounds grow linearly, which is in contrast with the general bound  $\bar{\epsilon}$  in Proposition 1 that decays exponentially and is known to be conservative. This is due to that, from the robust control viewpoint, the uncertainty of  $\mu$  in the polynomial coefficients becomes smaller as  $n$  increases.

Alternatively, we employ the Jury stability test [34] to derive that the zeros of the polynomial  $p(\lambda)$  are strictly inside the unit circle if and only if

$$\beta_0 := \begin{vmatrix} 1 & \alpha_0 \\ \bar{\alpha}_0 & 1 \end{vmatrix} > 0, \quad \beta_1 := \begin{vmatrix} \begin{vmatrix} 1 & \alpha_0 \\ \bar{\alpha}_0 & 1 \end{vmatrix} & \begin{vmatrix} 1 & \alpha_1 \\ \bar{\alpha}_0 & \bar{\alpha}_1 \end{vmatrix} \\ \begin{vmatrix} 1 & \bar{\alpha}_1 \\ \alpha_0 & \alpha_1 \end{vmatrix} & \begin{vmatrix} 1 & \bar{\alpha}_0 \\ \alpha_0 & 1 \end{vmatrix} \end{vmatrix} > 0. \quad (19)$$

Here  $\beta_0$  and  $\beta_1$  turn out to be polynomials in  $\epsilon$  of second and fourth order, respectively; the corresponding coefficients are functions of  $\mu$  and  $n$ . Thus selecting  $\mu$  such that  $\mu \neq 0$  and  $|\mu - 1/(2n)| \leq 1/(2n)$ , we

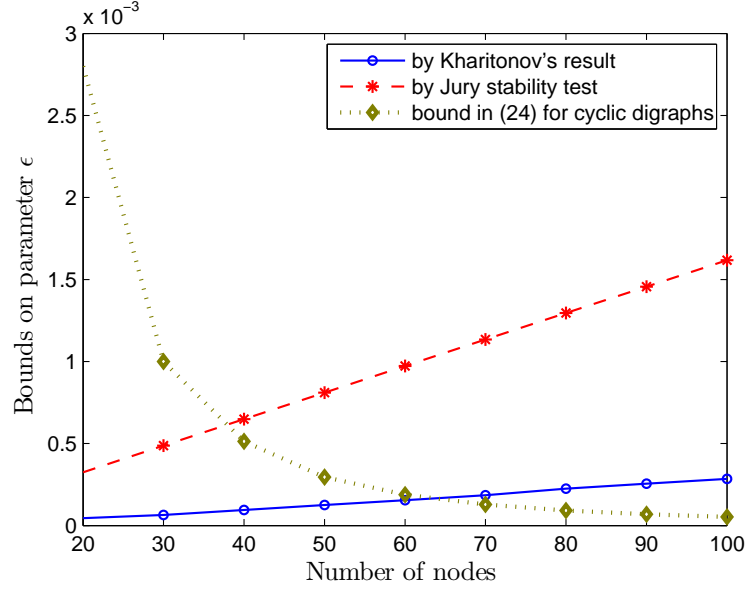


Fig. 2. Upper bounds on parameter  $\epsilon$  such that deterministic algorithm (3) achieves average consensus on general strongly connected balanced digraphs (solid and dashed curves) and cyclic digraphs (dotted curve).

can solve the inequalities in (19) for  $\epsilon$  in terms of  $n$ . Thereby we obtain the dashed curve in Fig. 2, each plotted point being the minimum value of  $\epsilon$  over 1000 random samples such that the inequalities in (19) hold. This simulation confirms that the true bound on  $\epsilon$  for the zeros of  $p(\lambda)$  to be inside the unit circle is between the solid and dashed curves. Since the discrepancy of these two curves is relatively small, it is suggested that our previous analysis based on Kharitonov's result may not very conservative.

Here ends our discussion on  $\epsilon$  bounds for arbitrary balanced (and strongly connected) digraphs. In the sequel, we will further specialize the balanced digraph  $\mathcal{G}$  to be symmetric or cyclic, respectively, and provide analytic  $\epsilon$  bounds less conservative than (5) for the general case. In particular, the exponent  $n$  is not involved.

#### A. Connected Undirected Graphs

A digraph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  is *symmetric* if  $(j, i) \in \mathcal{E}$  implies  $(i, j) \in \mathcal{E}$ . That is,  $\mathcal{G}$  is undirected.

*Proposition 4.* Consider a general connected undirected graph  $\mathcal{G}$ . Then the deterministic algorithm (3) achieves average consensus if the parameter  $\epsilon$  satisfies  $\epsilon \in (0, (1 - (1/n))(2 - (1/n)))$ .

*Proof.* The symmetry of the undirected graph  $\mathcal{G}$  results in the symmetry of its Laplacian matrix  $L$ . So all the eigenvalues  $\mu_i$  of  $L$  are real, and satisfy  $\mu_1 = 0$  and  $(\forall i \in [2, n]) \mu_i \in (0, 1/n]$  ( $\mathcal{G}$  is connected). For

$\mu_1 = 0$  we know from (17) that  $\lambda_1 = 1$ , and  $\lambda_2 \in (-1, 1)$  since  $0 < \epsilon < (1 - (1/n))(2 - (1/n)) < 2$ . For  $\mu_i \in (0, 1/n]$ ,  $i \in [2, n]$ , consider again the polynomial  $p(\lambda)$  in (16). According to the Jury stability test for real-coefficient case [35], the zeros of  $p(\lambda)$  are strictly inside the unit circle if and only if

$$1 + \alpha_0 + \alpha_1 > 0, \quad 1 + \alpha_0 - \alpha_1 > 0, \quad |\alpha_0| < 1.$$

Straightforward calculations show that these conditions hold provided  $\epsilon \in (0, (1 - (1/n))(2 - (1/n)))$ . Hence, the matrix  $M$  has a simple eigenvalue  $\lambda_1 = 1$  and all others  $\lambda_2, \dots, \lambda_{2n} \in (0, 1)$ . Therefore, by Proposition 2 the deterministic algorithm (3) achieves average consensus. ■

It is noted that for connected undirected graphs, the upper bound on  $\epsilon$  ensuring average consensus grows as  $n$  increases. This characteristic is in agreement with that of the bounds for the more general class of balanced digraphs as we observed in Fig. 2.

### B. Cyclic Digraphs

A digraph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  is *cyclic* if  $\mathcal{V} = \{1, \dots, n\}$  and  $\mathcal{E} = \{(1, 2), (2, 3), \dots, (n-1, n), (n, 1)\}$ . So a cyclic digraph is strongly connected.

*Proposition 5.* Suppose that the digraph  $\mathcal{G}$  is cyclic. Then the deterministic algorithm (3) achieves average consensus if the parameter  $\epsilon$  satisfies

$$\epsilon \in \left(0, \frac{\sqrt{2}}{3 + \sqrt{5}}(1 - |\lambda_3|)\right), \quad \text{with } \lambda_3 \text{ as in (8)}. \quad (20)$$

Further, in this case  $|\lambda_3| = \sqrt{1 - (1/n) + (1/(2n^2)) + (1/n)(1 - 1/(2n)) \cos 2\pi/n}$ .

Before providing the proof, we state a perturbation result, the Bauer–Fike Theorem, for diagonalizable matrices (e.g., [28, Section 6.3]). Recall that the matrix  $M$  in (3) can be written as  $M = M_0 + \epsilon F$ , with  $M_0$  and  $F$  in (7). Throughout this subsection, write  $\lambda_i(\epsilon)$  for the eigenvalues of  $M$ , and  $\lambda_i$  for those of  $M_0$ .

*Lemma 8.* Suppose that  $M_0$  is diagonalizable; i.e., there exist a nonsingular matrix  $V \in \mathbb{C}^{2n \times 2n}$  and a diagonal matrix  $J = \text{diag}(\lambda_1, \dots, \lambda_{2n})$  such that  $M_0 = VJV^{-1}$ . If  $\lambda(\epsilon)$  is an eigenvalue of  $M$ , then there is an eigenvalue  $\lambda_i$  of  $M_0$ , for some  $i \in [1, 2n]$ , such that  $|\lambda(\epsilon) - \lambda_i| \leq \|V\|_2 \|V^{-1}\|_2 \|\epsilon F\|_2$ .

In other words, every eigenvalue of the perturbed matrix  $M$  lies in a circle centered at some eigenvalue of the unperturbed matrix  $M_0$  of the radius  $(\|V\|_2 \|V^{-1}\|_2 \|\epsilon F\|_2)$ . We now present the proof of Proposition 5.

*Proof of Proposition 5.* Since the digraph  $\mathcal{G}$  is cyclic, we derive that its Laplacian matrix  $L$  is given by  $L = \text{circ}(1/(2n), 0, \dots, 0, -1/(2n))$  – a *circulant matrix* [36] with the first row  $[1/(2n) \ 0 \ \dots \ 0 \ -1/(2n)] \in$

$\mathbb{R}^{1 \times n}$ . Let  $\omega := e^{2\pi\iota/n}$  with  $\iota := \sqrt{-1}$ . Then the eigenvalues  $\mu_i$  of  $L$  are  $\mu_i = (1/(2n))(1 - \omega^{i-1})$ ,  $i = 1, \dots, n$ . Rewrite the equation (17) as  $(\lambda(\epsilon) - 1)(\lambda(\epsilon) - 1 + \epsilon) + 3(\lambda(\epsilon) - 1)\mu_i + 2\mu_i^2 = 0$ . Then for  $\mu_1 = 0$ , we have  $\lambda_1(\epsilon) = 1$  and  $\lambda_2(\epsilon) = 1 - \epsilon$ , corresponding respectively to the eigenvalues  $\lambda_1, \lambda_2$  of  $M_0$ . Evidently the upper bound in (20) is strictly smaller than 2; so  $\lambda_2(\epsilon) \in (-1, 1)$ .

We turn next to investigating the rest of the eigenvalues  $\lambda_3(\epsilon), \dots, \lambda_{2n}(\epsilon)$ , for which we employ Lemma 8. Let  $\Omega$  denote the  $n \times n$  *Fourier matrix* given by

$$\Omega := \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega & \omega^2 & \dots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \dots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \dots & \omega^{(n-1)(n-1)} \end{bmatrix}.$$

Note that  $\Omega$  is unitary, i.e.,  $\Omega^{-1} = \Omega^*$  (the conjugate transpose of  $\Omega$ ). It is a fact that every circulant matrix can be (unitarily) diagonalized by  $\Omega$  [36, Theorem 3.2.1]. Now let  $V := \begin{bmatrix} \Omega & 0 \\ \Omega & \Omega \end{bmatrix}$ , and consider

$$V^{-1}M_0V = \begin{bmatrix} \Omega^* & 0 \\ -\Omega^* & \Omega^* \end{bmatrix} \begin{bmatrix} I - L & 0 \\ L & S \end{bmatrix} \begin{bmatrix} \Omega & 0 \\ \Omega & \Omega \end{bmatrix} = \begin{bmatrix} \Omega^*(I - L)\Omega & 0 \\ 0 & \Omega^*S\Omega \end{bmatrix}.$$

The last equality is due to  $S = I - 2L$ . Hence  $M_0$  is diagonalizable via  $V$ , and its spectrum is

$$\sigma(M_0) = \sigma(I - L) \cup \sigma(S) = \left\{ 1 - \frac{1}{2n}(1 - \omega^{i-1}), 1 - \frac{1}{n}(1 - \omega^{i-1}) : i = 1, \dots, n \right\}.$$

Also, by a direct calculation we get  $\|V\|_2 = \|V^{-1}\|_2 = \sqrt{(3 + \sqrt{5})/2}$  and  $\|F\|_2 = \sqrt{2}$ . It then follows from Lemma 8 that for every eigenvalue  $\lambda_l(\epsilon)$  of  $M$  there is an eigenvalue  $\lambda_{l'}$  of  $M_0$ ,  $l, l' \in [3, 2n]$ , such that  $|\lambda_l(\epsilon) - \lambda_{l'}| \leq \|V\|_2 \|V^{-1}\|_2 \|\epsilon F\|_2 = ((3 + \sqrt{5})/2) \sqrt{2} \epsilon$ . So the upper bound of  $\epsilon$  in (20) guarantees  $|\lambda_l(\epsilon) - \lambda_{l'}| < 1 - |\lambda_3|$ ; namely, the perturbed eigenvalues still lie within the unit circle. Summarizing the above we have  $\lambda_1(\epsilon) = 1$  and  $|\lambda_2(\epsilon)|, |\lambda_3(\epsilon)|, \dots, |\lambda_{2n}(\epsilon)| < 1$ ; therefore, the deterministic algorithm (3) achieves average consensus by Proposition 2. Further, one computes that

$$\begin{aligned} |\lambda_3| &= \max_{i \in [2, n]} \left\{ \left| 1 - \frac{1}{2n}(1 - \omega^{i-1}) \right|, \left| 1 - \frac{1}{n}(1 - \omega^{i-1}) \right| \right\} \\ &= \left| 1 - \frac{1}{2n} + \frac{1}{2n}\omega \right| = \sqrt{1 - \frac{1}{n} + \frac{1}{2n^2} + \frac{1}{n} \left( 1 - \frac{1}{2n} \right) \cos \frac{2\pi}{n}}. \end{aligned}$$

■

Finally, in Fig. 2 we plot the upper bound on  $\epsilon$  in (20) for the class of cyclic digraphs. We see that this bound decays as the number  $n$  of nodes increases, which contrasts with the bound characteristic

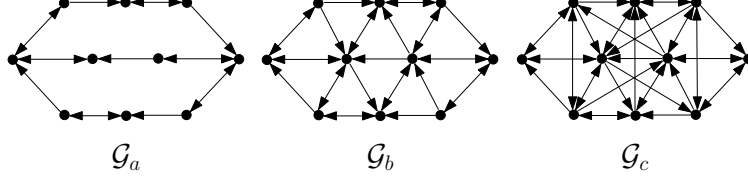


Fig. 3. Three examples of strongly connected but non-balanced digraphs.

TABLE I  
CONVERGENCE FACTORS  $\lambda_2^{(d)}$  AND  $\lambda_2^{(g)}$  WITH RESPECT TO DIFFERENT VALUES OF PARAMETER  $\epsilon$ .

	$\epsilon = 0.2$		$\epsilon = 0.7$		$\epsilon = 2.15$	
	$\lambda_2^{(d)}$	$\lambda_2^{(g)}$	$\lambda_2^{(d)}$	$\lambda_2^{(g)}$	$\lambda_2^{(d)}$	$\lambda_2^{(g)}$
$\mathcal{G}_a$	0.9963	0.9963	0.9993	1.0003	1.0003	1.0020
$\mathcal{G}_b$	0.9951	0.9951	0.9969	0.9969	0.9985	1.0000
$\mathcal{G}_c$	0.9883	0.9883	0.9930	0.9930	0.9966	0.9993

of the more general class of balanced digraphs. This may indicate the conservativeness of our current approach based on perturbation theory. Nevertheless, since the perturbation result used here is specific only to diagonalizable matrices, the derived upper bound in (20) is less conservative than the general one in (5).

## VI. NUMERICAL EXAMPLES

### A. Convergence Paths

Consider the three digraphs displayed in Fig. 3, with 10 nodes and respectively 17, 29, and 38 edges. Note that all the digraphs are strongly connected, and in the case of uniform weights they are non-balanced (indeed, no single node is balanced). We apply both the deterministic algorithm (3), with uniform weights  $a = 1/(2\text{card}(\mathcal{E}))$  and  $b = 1/\text{card}(\mathcal{E})$ , and the gossip algorithm (12), with uniform weight  $w = 1/2$  and probability  $p = 1/\text{card}(\mathcal{E})$ .

The convergence factors  $\lambda_2^{(d)}$  and  $\lambda_2^{(g)}$  (see Remarks 1 and 3) for three different values of the parameter  $\epsilon$  are summarized in Table I. We see that small  $\epsilon$  ensures convergence of both algorithms (the gossip algorithm (12) requires smaller values of  $\epsilon$  for mean-square convergence), whereas large  $\epsilon$  can lead to instability. Moreover, in those converging cases the factors  $\lambda_2^{(d)}$  and  $\lambda_2^{(g)}$  decrease as the number of edges increases from  $\mathcal{G}_a$  to  $\mathcal{G}_c$ , which indicates faster convergence when there are more communication channels

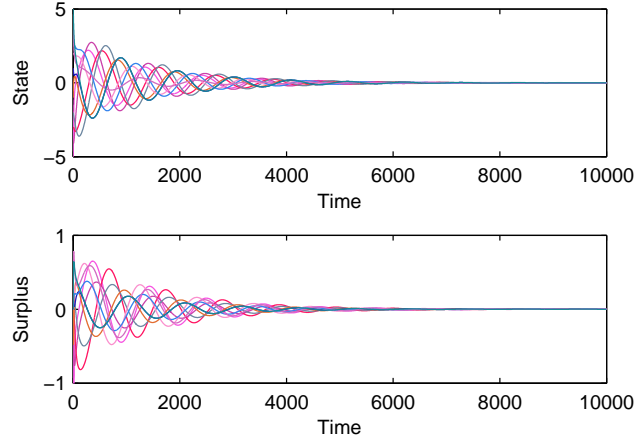


Fig. 4. Convergence paths of states and surpluses: Obtained by applying the deterministic algorithm (3) with parameter  $\epsilon = 0.7$  on digraph  $\mathcal{G}_a$ .

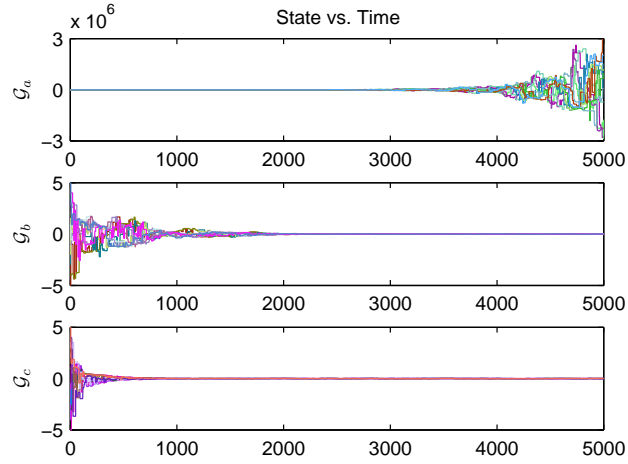


Fig. 5. Sample paths of states: Obtained by applying the gossip algorithm (3) with parameter  $\epsilon = 0.7$  on digraphs  $\mathcal{G}_a$ ,  $\mathcal{G}_b$ , and  $\mathcal{G}_c$ .

available for information exchange. We also see that the algorithms are more robust on digraphs with more edges, in the sense that a larger range of values of  $\epsilon$  is allowed.

For a random initial state  $x(0)$  with the average  $x_a = 0$  and the initial surplus  $s(0) = 0$ , we display in Fig. 4 the trajectories of both states and surpluses when the deterministic algorithm (3) is applied on digraph  $\mathcal{G}_a$  with parameter  $\epsilon = 0.7$ . Observe that asymptotically, state averaging is achieved and surplus vanishes. Under the same conditions, the gossip algorithm (12), however, fails to converge, as shown in Fig. 5. Applying algorithm (12) instead on the digraphs  $\mathcal{G}_b$  and  $\mathcal{G}_c$  which have more edges, average consensus is again reached, and faster convergence occurs in  $\mathcal{G}_c$  (see Fig. 5).

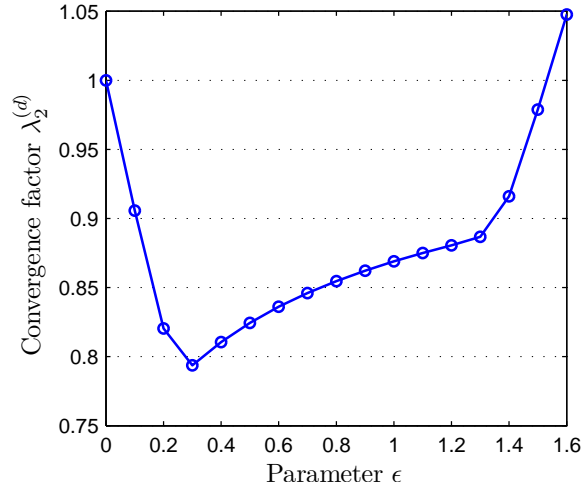


Fig. 6. Convergence factor  $\lambda_2^{(d)}$  of the deterministic algorithm (3) with respect to parameter  $\epsilon$ .

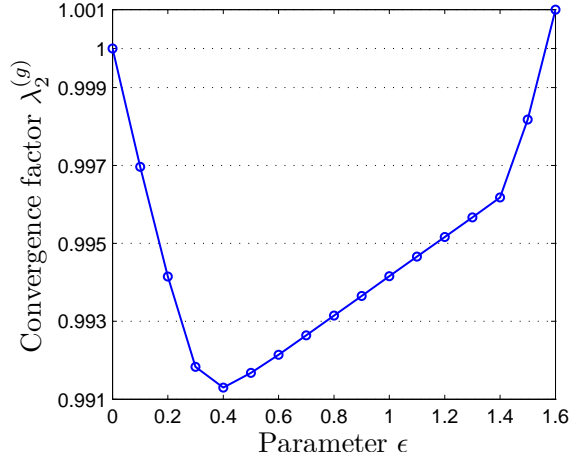


Fig. 7. Convergence factor  $\lambda_2^{(g)}$  of the gossip algorithm (12) with respect to parameter  $\epsilon$ .

### B. Convergence Speed versus Parameter $\epsilon$

We have seen that a sufficiently small parameter  $\epsilon$  ensures convergence of both algorithms (3) and (12). Now we investigate the influence of  $\epsilon$  on the speed of convergence, specifically the convergence factors  $\lambda_2^{(d)}$  and  $\lambda_2^{(g)}$ . To reduce the effect of network topology in this investigation, we employ a type of random digraphs where an edge between every pair of nodes can exist with probability  $1/2$ , independent across the network and invariant over time; we take only those that are strongly connected.

For the deterministic algorithm (3), consider random digraphs of 50 nodes and uniform weights  $a = b = 1/50$ . Fig. 6 displays the curve of convergence factor  $\lambda_2^{(d)}$  with respect to the parameter  $\epsilon$ , each

plotted point being the mean value of  $\lambda_2^{(d)}$  over 100 random digraphs. To account for the trend of this curve, first recall from the perturbation argument for Theorem 1 that the matrix  $M$  in (3) has two (maximum) eigenvalues 1 when  $\epsilon = 0$ , and small  $\epsilon$  causes that one of them (denote its modulus by  $\lambda_{in}$ ) moves into the unit circle. Meanwhile, some other eigenvalues of  $M$  inside the unit circle move outward; denote the maximum modulus among these by  $\lambda_{out}$ . In our simulations it is observed that when  $\epsilon$  is small,  $\lambda_2^{(d)} = \lambda_{in}$  ( $> \lambda_{out}$ ) and  $\lambda_{in}$  moves further inside as perturbation becomes larger; so  $\lambda_2^{(d)}$  decreases (faster convergence) as  $\epsilon$  increases in the beginning. Since the eigenvalues move continuously, there exists some  $\epsilon$  such that  $\lambda_{in} = \lambda_{out}$ , corresponding to the fastest convergence speed. After that,  $\lambda_2^{(d)} = \lambda_{out}$  ( $> \lambda_{in}$ ) and  $\lambda_{out}$  moves further outside as  $\epsilon$  increases; hence  $\lambda_2^{(d)}$  increases and convergence becomes slower, and finally divergence occurs.

An analogous experiment is conducted for the gossip algorithm (12), with random digraphs of 30 nodes, uniform probability  $p = 1/\text{card}(\mathcal{E})$ , and uniform weight  $w_{ij} = 1/2$ . We see in Fig. 7 a similar trend of  $\lambda_2^{(g)}$  as the parameter  $\epsilon$  increases, though it should be noted that the changes in  $\lambda_2^{(g)}$  are smaller than those in  $\lambda_2^{(d)}$ . From these observations, it would be of ample interest to exploit the values of  $\epsilon$  when the convergence factors achieve their minima, as well as the upper bounds of  $\epsilon$  ensuring convergence.

## VII. CONCLUSIONS

We have proposed distributed algorithms which enable networks of agents to achieve average consensus on arbitrary strongly connected digraphs. Specifically, in synchronous networks a deterministic algorithm ensures asymptotic state averaging, and in asynchronous networks a gossip algorithm guarantees average consensus in the mean-square sense and with probability one. To emphasize, our derived graphical condition is more general than those previously reported in the literature, in the sense that it does not require balanced network structure; also, the matrix perturbation theory plays an important role in the convergence analysis. Moreover, special regular digraphs are investigated to give less conservative bounds on the parameter  $\epsilon$ ; and numerical examples are provided to illustrate the convergence results, with emphasis on convergence speed.

For future research, one direction of interest would be to extend the deterministic algorithm (3) to the more realistic scenario of *switching* digraphs  $\mathcal{G}(k) = (\mathcal{V}, \mathcal{E}(k))$ ; namely, the network topology is time-varying. If every  $\mathcal{G}(k)$ ,  $k \geq 0$ , is strongly connected, then it is possible to ensure convergence by introducing *slow switching* (e.g., dwell time) as in [5], [37]. Under the weaker graphical condition that digraphs  $\mathcal{G}(k)$  are *jointly* strongly connected ([2], [27]), to verify if average consensus can be achieved seems to be more challenging and requires further investigation.

On the other hand, in the literature on gossip algorithms [6], [7], [38], a variety of practical communication issues have been discussed such as link failure, message collision, broadcast protocol, and synchronized node selection (i.e., multiple nodes are selected at the same time). We thus aim at addressing these issues by making suitable extensions of our gossip algorithm (12).

#### ACKNOWLEDGMENT

The authors would like to thank Sandro Zampieri for the helpful discussion, and the anonymous reviewers for the valuable comments.

#### REFERENCES

- [1] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Prentice-Hall, 1989.
- [2] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom. Control*, vol. 48, no. 6, pp. 988–1001, 2003.
- [3] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520–1533, 2004.
- [4] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, no. 1, pp. 65–78, 2004.
- [5] W. Ren and R. W. Beard, *Distributed Consensus in Multi-vehicle Cooperative Control: Theory and Applications*. Springer-Verlag, 2008.
- [6] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.
- [7] A. Kashyap, T. Başar, and R. Srikant, "Quantized consensus," *Automatica*, vol. 43, no. 7, pp. 1192–1203, 2007.
- [8] R. Carli, F. Fagnani, P. Frasca, and S. Zampieri, "Gossip consensus algorithms via quantized communication," *Automatica*, vol. 46, no. 1, pp. 70–80, 2010.
- [9] J. Lavaei and R. M. Murray, "Quantized consensus by means of gossip algorithm," *IEEE Trans. Autom. Control*, vol. 57, no. 1, pp. 19–32, 2012.
- [10] H. Ishii and R. Tempo, "Distributed randomized algorithms for the PageRank computation," *IEEE Trans. Autom. Control*, vol. 55, no. 9, pp. 1987–2002, 2010.
- [11] K. Cai and H. Ishii, "Average consensus on general digraphs," in *Proc. 50th IEEE Conf. on Decision and Control and Euro. Control Conf.*, Orlando, FL, 2011, pp. 1956–1961.
- [12] M. Franceschelli, A. Giua, and C. Seatzu, "Consensus on the average on arbitrary strongly connected digraphs based on broadcast gossip algorithms," in *Proc. 1st IFAC Workshop on Estimation and Control of Networked Systems*, Venice, Italy, 2009, pp. 66–71.
- [13] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Proc. 44th Annual IEEE Symposium on Foundations of Computer Science*, Washington DC, 2003, pp. 482–491.
- [14] F. Benezit, V. Blondel, P. Thiran, J. Tsitsiklis, and M. Vetterli, "Weighted gossip: distributed averaging using non-doubly stochastic matrices," in *Proc. IEEE Int. Symposium on Information Theory*, Austin, TX, 2010, pp. 1753–1757.

- [15] B. Gharesifard and J. Cortés, “Distributed strategies for generating weight-balanced and doubly stochastic digraphs,” *arXiv:0911.0232*, 2011.
- [16] S. Patterson, B. Bamieh, and A. El Abbadi, “Distributed average consensus with stochastic communication failures,” in *Proc. 46th IEEE Conf. on Decision and Control*, New Orleans, LA, 2007, pp. 4215–4220.
- [17] F. Fagnani and S. Zampieri, “Average consensus with packet drop communication,” *SIAM J. Control Optimization*, vol. 48, no. 1, pp. 102–133, 2009.
- [18] T. Aysal, B. Oreshkin, and M. Coates, “Accelerated distributed average consensus via localized node state prediction,” *IEEE Trans. Signal Processing*, vol. 57, no. 4, pp. 1563–1576, 2009.
- [19] J. Liu, B. D. O. Anderson, M. Cao, and A. S. Morse, “Analysis of accelerated gossip algorithms,” in *Proc. 48th IEEE Conf. on Decision and Control*, Shanghai, China, 2009, pp. 871–876.
- [20] A. P. Seyranian and A. A. Mailybaev, *Multiparameter Stability Theory with Mechanical Applications*. World Scientific, 2004.
- [21] R. Bhatia, *Matrix Analysis*. Springer-Verlag, 1996.
- [22] G. W. Stewart and J. Sun, *Matrix Perturbation Theory*. Academic Press, 1990.
- [23] K. Cai and H. Ishii, “Quantized consensus and averaging on gossip digraphs,” *IEEE Trans. Autom. Control*, vol. 56, no. 9, pp. 2087–2100, 2011.
- [24] G. Grimmett and D. Stirzaker, *Probability and Random Processes*. Oxford University Press, 2001.
- [25] Z. Li, Z. Duan, G. Chen, and L. Huang, “Consensus of multiagent systems and synchronization of complex networks: a unified viewpoint,” *IEEE Trans. Circuits and Systems*, vol. 57, no. 1, pp. 213–224, 2010.
- [26] T. Li, M. Fu, L. Xie, and J. Zhang, “Distributed consensus with limited communication data rate,” *IEEE Trans. Autom. Control*, vol. 56, no. 2, pp. 279–292, 2011.
- [27] Z. Lin, *Distributed Control and Analysis of Coupled Cell Systems*. VDM Verlag, 2008.
- [28] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.
- [29] Y. Cao and W. Ren, “Multi-vehicle coordination for double-integrator dynamics under fixed undirected/directed interaction in a sampled-data setting,” *Int. J. of Robust and Nonlinear Control*, vol. 20, no. 9, pp. 987–1000, 2010.
- [30] F. R. Gantmacher, *The Theory of Matrices, Vol. 2*. Chelsea, 1959.
- [31] O. Costa, M. Fragoso, and R. Marques, *Discrete-Time Markov Jump Linear Systems*. Springer-Verlag, 2004.
- [32] J. R. Silvester, “Determinants of block matrices,” *The Mathematical Gazette*, vol. 84, no. 501, pp. 460–467, 2000.
- [33] B. R. Barmish, *New Tools for Robustness of Linear Systems*. Macmillan, 1994.
- [34] E. I. Jury, “Modified stability table for 2-d digital filters,” *IEEE Trans. Circuits and Systems*, vol. 35, no. 1, pp. 116–119, 1988.
- [35] —, “A modified stability table for linear discrete systems,” *Proc. IEEE*, vol. 53, no. 2, pp. 184–185, 1965.
- [36] P. J. Davis, *Circulant Matrices*. AMS Chelsea Publishing, 1994.
- [37] D. W. Casbeer, R. Beard, and A. L. Swindlehurst, “Discrete double integrator consensus,” in *Proc. 47th IEEE Conf. on Decision and Control*, Cancun, Mexico, 2008, pp. 2264–2269.
- [38] P. Frasca and F. Fagnani, “Broadcast gossip averaging algorithms: Interference and asymptotical error in large networks,” *arXiv: 1005.1292*, 2010.